Metallomics reveals a persisting impact of cadmium on the evolution of metal-selective snail metallothioneins

Dallinger, Reinhard ; Zerbe, Oliver ; Baumann, Christian ; Egger, Bernd ; Capdevila, Merce ; Palacios, Oscar ; Albalat, Ricard ; Calatayud, Sara ; Ladurner, Peter ; Schlick-Steiner, Birgit C ; Steiner, Florian M ; Pedrini-Martha, Veronica ; Lackner, R ; Lindner, Herbert ; Dvorak, M ; Niederwanger, Michael

Abstract: The tiny contribution of cadmium (Cd) to the composition of the earth's crust contrasts with its high biological significance, owing mainly to the competition of Cd with the essential zinc (Zn) for suitable metal binding sites in proteins. In this context it was speculated that in several animal lineages, the protein family of metallothioneins (MTs) has evolved to specifically detoxify Cd. Although the multi-functionality and heterometallic composition of MTs in most animal species does not support such an assumption, there are some exceptions to this role, particularly in animal lineages at the roots of animal evolution. In order to substantiate this hypothesis and to further understand MT evolution, we have studied MTs of different snails that exhibit clear Cd-binding preferences in a lineage-specific manner. By applying a metallomics approach including 74 MT sequences from 47 gastropod species, and by combining phylogenomic methods with molecular, biochemical, and spectroscopic techniques, we show that Cd selectivity of snail MTs has resulted from convergent evolution of metal-binding domains that significantly differ in their primary structure. We also demonstrate how their Cd selectivity and specificity has been optimized by the persistent impact of Cd through 430 million years of MT evolution, modifying them upon lineage-specific adaptation of snails to different habitats. Overall, our results support the role of Cd for MT evolution in snails, and provide an interesting example of a vestigial abiotic factor directly driving gene evolution. Finally, we discuss the potential implications of our findings for studies devoted to the understanding of mechanisms leading to metal specificity in proteins, which is important when designing metal-selective peptides.

DOI: https://doi.org/10.1039/c9mt00259f

Metallomics reveal a persisting impact of cadmium on the evolution of metal-selective snail metallothioneins

by

Reinhard Dallinger[1,2*], Oliver Zerbe[3*], Christian Baumann[3], Bernhard Egger[1], Mercé Capdevila[4], Òscar Palacios[4], Ricard Albalat[5], Sara Calatayud [5], Peter Ladurner[1,2], Birgit Schlick-Steiner[6], Florian Steiner[6], Veronika Pedrini-Martha[1], Reinhard Lackner[1], Herbert Lindner[7], Martin Dvorak[1], Michael Niederwanger[1], Raimund Schnegg[1], Silvia Atrian[5,†]

[1] Department of Zoology, University of Innsbruck, Austria
[2] Center for Molecular Biosciences Innsbruck, Austria
[3] Department of Chemistry, University of Zürich, Switzerland
[4] Departament de Química, Universitat Autònoma de Barcelona, Spain
[5] Deapartment of Genetics, University of Barcelona, Spain
[6] Department of Ecology, University of Innsbruck, Austria
[7] Division of Clinical Biochemistry, Innsbruck Medical University, Austria
[†] deceased. We want to dedicate this work to our longtime cooperation partner, Silvia Atrian

[*] Corresponding authors: reinhard.dallinger@uibk.ac.at and oliver.zerbe@chem.uzh.ch

Abstract

The tiny contribution of cadmium (Cd) to the composition of the earth crust contrasts with its high biological significance, owing mainly to the competition of Cd with the essential zinc (Zn) for suitable metal binding sites in proteins. In this context it was speculated that in several animal lineages, the protein family of metallothioneins (MTs) has evolved to specifically detoxify Cd. Although the multi-functionality and heterometallic composition of MTs in most animal species does not support such an assumption, there are some exceptions from this role, particularly in animal lineages at the roots of animal evolution. In order to substantiate this hypothesis and to further understand MT evolution, we have studied MTs of different snails that exhibit clear Cd-binding preferences in a lineage-specific manner. By applying a metallomics approach including 74 MT sequences from 47 gastropod species, and by combining phylogenomic methods with molecular, biochemical, and spectroscopic techniques, we show that Cd selectivity of snail MTs has resulted from convergent evolution of metal-binding domains that significantly differ in their primary structure. We also demonstrate how their Cd selectivity and specificity has been optimized by the persistent impact of Cd through 430 million years of MT evolution, modifying them upon lineage-specific adaptation of snails to different habitats. Overall, our results support the role of Cd for MT evolution in snails, and provide an interesting example of a vestigial abiotic factor directly driving gene evolution. Finally, we discuss the potential implications of our findings for studies devoted to the understanding of mechanisms leading to metal specificity in proteins, which is important when designing metal-selective peptides.

Introduction

With a tiny amount of about 0.00001%, the contribution of cadmium (Cd) to the composition of the earth crust is seemingly negligible [1]. In spite of this, the biological significance of Cd is distinctly higher, owing to its particular patchy distribution, enrichment and circulation in the biosphere [2]. In some diatom marine algae, for example, Cd has achieved an essential importance as a constituent of the algal enzyme carbonic anhydrase, a fact that has been explained by the relative preponderance of Cd at the cost of lowly available Zn in oceanic environments inhabited by these algae [3,4]. In most organisms, however, Cd is highly toxic at very low concentrations, due to its physico-chemical similarity and competition with zinc (Zn) [5], one of the most important essential trace elements. Because of this, most organisms have developed strategies for Cd handling and detoxification [6], and it has been hypothesized that metallothioneins (MTs), a ubiquitous protein family with a high affinity to transition metal ions, may have been developed by organisms to clear this highly toxic metal [7]. Yet, this hypothesis has been questioned because of the apparent involvement of most MTs in a variety of functions, and their often heterometallic and metamorphic composition with binding affinities to different metal ions [8–10]. However, MTs form a huge and diverse gene superfamily present in most kingdoms of organisms, from bacteria through fungi, plants and animals [11,12]. This suggests that their origins may go back to the primal evolutionary roots of life on earth, although the metal preference of the ancestral MT remains unknown. In contrast to modern vertebrates some MTs at the roots of, for example, Chordata are Cd-selective, as recently reported for MTs of the tunicate *Oikopleura dioica* [13]. Cd- and Cu-selective MTs have also been discovered in several species of the ancient mollusk class of Gastropoda (snails and slugs) [14,15]. This suggests that in early evolution of life, Cd-selectivity of MTs might have been more common than today, and this feature has evidently been preserved to the present in diverse animal clades while it disappeared in others.

To support this hypothesis, we have taken advantage of the Cd-specific gastropod MTs, which provide an ideal model system to study the evolutionary influence of Cd on MT evolution along more of 400 million years (MY) of Gastropoda diversification. Unlike many other modern animals, snails possess metal-selective MTs, such as Cd- (CdMTs) and Cu-selective (CuMTs) isoforms [15] that perform Cd- or Cu-specific tasks. Thereby they exhibit a straightforward relationship between metal binding features and related physiological functions. Interestingly, Cd-specific snail MTs bind this metal with a strength and exclusive preference hardly observed in any other protein family. They are expressed in a multitude of isoforms that vary in a clade-specific manner allowing us to compare and evaluate similar proteins and protein variants (and their metal-binding modifications) in a large number of species that have adapted to different habitats. These spread from marine through terrestrial to freshwater environments with significantly different Cd concentrations. This comparative approach is central to understand how MTs have been optimized for Cd binding during gastropod evolution by the continuous impact of Cd, and how its influence is modulated by habitat-specific environmental constraints.

In our work, we applied a comprehensive metallomics approach by characterizing 74 novel or known MT sequences from 47 species across all major gastropod clades [14–23]. We used phylogenomic methods based on next generation sequencing to obtain transcriptomic data for evolutionary analyses and construction of phylogenetic trees. We also analyzed neutral DNA markers to compare the resulting phylogenetic tree with MT-derived trees. In addition, we provide data on metal-selective features of recombinant snail MTs and their metal-binding domains, based on molecular, biochemical and spectroscopic methods. Our data indicate that Cd selectivity has evolved since 430 million years ago (MYA) in gastropod MTs through convergent evolution of metal-binding domains with diverging primary structures. We study the mechanisms by which their Cd binding features have been optimized, and illustrate how they have diversified into different kinds with altered or even lost metal selectivity through lineage-specific transition into novel habitats that differ in their natural Cd background concentrations. Overall, we have been able to demonstrate a continuous impact of Cd on evolution of one of the most important metal-binding protein families, and describe a paradigmatic case of how an abiotic factor directly drives gene evolution. Finally, we discuss possible implications of our findings to better understand how metal-selectivity has been achieved in nature, and how this knowledge can help in designing metal-selectivity in synthetic peptides.

Material and methods

Animal collection, rearing and Cd exposure

A list of gastropod species involved in experimental work for the present study along with methodical applications is reported in **Table 1**.

Individuals of *Alinda biplicata* and *Deroceras reticulatum* were collected in suburbs of Innsbruck (Tyrol, Austria) in 2017 and 2018. Individuals of *Patella vulgata* were collected in Barcelona, Spain in summer 2016. Snails of the helicid species *Cornu aspersum* were bought from a commercial dealer (Wiener Schnecken Manufaktur, Vienna, Austria), as were the aquatic species *Marisa cornuaretis*, *Anentome helena, Physa acuta* and *Aplysia californica* (Aquaristikzentrum Innsbruck, Tyrol, Austria). Adult individuals of *Lottia gigantea* were collected and kindly provided to us by Dr. Douglas J. Eernisse (California State University, Fullerton, Ca, USA).

For Cd exposure of *Cornu aspersum*, adult snails were acclimatized on garden earth substrate containing lime powder ($CaCO_3$) in groups of 30 individuals each under stable conditions in a climate chamber (18°C, 12h light/dark cycle) and were fed regularly with uncontaminated lettuce (*Lactuca sativa*) under moistened conditions for one week. For Cd exposure, control snails were fed with uncontaminated lettuce whereas Cd-exposed snails were fed with Cd-enriched lettuce which had been incubated for one

hour in a $CdCl_2$-solution containing 2 mg/l Cd^{2+} [24]. After five days of exposure, five individuals of each group were sacrificed.

Individuals of *Lymnaea stagnalis* were collected from an unpolluted freshwater pond in the Ternopil region, Ukraine (49o49/ N, 25o23/ E) and were kindly provided to us by Dr. Oksana B. Stoliar and Dr. Halina I. Falfushynska (Ternopil National Pedagogical University, Ukraine). For metal exposure, snails were kept in 80 l tanks of aerated tap water during 14 days, and exposed to a Cd concentration of 15 µg/l (in tap water). Water and Cd solutions were renewed every two days, lettuce feed was provided before water exchange. Control snails without Cd addition were kept in the same manner as exposed individuals. At the end of the exposure, three individuals per goup were dissected for mRNA extraction from midgut gland.

Dissection, RNA/DNA isolation and cDNA synthesis

Snails were sacrificed and midgut gland tissue of individual snails (n = 3-5) (*Patella vulgata*, *Lottia gigantea*, *Anentome helena*, *Marisa cornuarietis, Aplysia californica*, *Cornu aspersum*, *Deroceras reticulatum*, *Lymnaea stagnalis*) or – due to the small size of some species – mixed tissue parts (*Alinda biplicata*, *Physella acuta*) were dissected and stored in RNAlater® (Fisher Scientific, Vienna, Austria) at -80 °C. For quantitative Real Time PCR (qPCR) after metal exposure, small aliquots of midgut gland tissue (approx. 1 mg fresh weight) of control and Cd-exposed *Cornu aspersum* and *Lymnaea stagnalis* (n = 3-5) were transferred to RNAlater® (Fisher Scientific, Vienna, Austria) whereas the remaining part of the tissue was collected for metal measurement.

RNA tissue samples were homogenized with a Precellys® homogenizer (Bertin Instruments, Montigny-le-Bretonneux, France). Total RNA was isolated using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) including the on-column DNase I digestion according to the manufacturer's instructions (Qiagen). RNA integrity was checked by agarose gel electrophoresis and concentrations were estimated with Nanodrop (Thermo Fisher Scientific, Waltham, CA, USA). For qPCR, RNA samples were measured in triplicates with the Quant-iT™ Ribogreen® RNA Assay Kit (Life Technologies Corporation, Carlsbad, USA) applying the Victor™ X4 Multilable Reades (Perkin Elmer, Waltham, USA). 450 ng of total RNA was transcribed to cDNA in a 50 µl approach with the RevertAid Reverse Transcriptase (Thermo Fisher Scientific). For amplification of the multidomain MT sequences (*Alinda biplicata*, *Marisa cornuarietis*) AccuScript Hi-Fi Reverse transcriptase (Agilent, Santa Clara, CA, USA) was used in a 20 µl approach for cDNA synthesis.

For phylogenetic reconstruction based on neutral markers, DNA of the same specimens mentioned above was extracted using GenElute™ Mammalian Genomic DNA Miniprep Kit (Sigma Aldrich). A ~590 bp stretch of the mitochondrial gene Cytochrome C oxidase I (COI) was amplified using the standard

primers LCO1490 and HCO2198 suggested by [25] and the degenerated primers LoboF1 and LoboR1 [26] for some species.

PCR products of *Neritina pulligera* and *Littorina littorea* showed multiple bands and therefore were cloned before sequencing using the pTZ57R/T InsTAclone Kit (Thermo Fisher, Waltham, USA).

To obtain ~1000 bp of 18SrDNA the following primers of [25] were used in various combinations: 18A1, 470F, 1155F 700R, 1500R, 1800. PCR products were purified and Sanger sequenced by the facilities of Eurofins (MWG Operon, Germany). For *Helix pomatia* and *Patella vulgate*, only ~500 bp of 18SrDNA sequence were available; thus, full sequences were obtained from GenBank (FJ977750, FJ977632, AF239734, AY145373, MF544434, AY427527). Conditions for thermal cycling, polymerase and PCR are shown in **Table S1**, newly generated sequences have been deposited at GenBank (MK919674-MK919701).

RNA seq and transcriptome assembly

Isolated RNA from an individual midgut gland (*Patella vulgata*, *Neritina pulligera*, *Littorina littorea*, *Pomatias elegans*, *Pomacea bridgesii*, *Marisa cornuarietis*, *Anentome helena*, *Elysia crispata*, and *Limax maximus*) or of pooled soft-tissue (*Alinda biplicata*) was sent to the Duke Center for Genomic and Computational Biology (GBC, Duke University, Durham, NC, USA) and sequenced with Hi-Seq 4000 Illumina sequencing. A separate library was sequenced for each species. Raw data were assembled using Trinity [27] version: v2.1.1 with default settings. Assemblies were provided for analysis on a local BLAST server "SequenceServer" [28], where cDNA sequences encoding for diverse snail MTs were blasted against the transcriptomic data sets to identify MT sequences. Raw sequence reads data were deposited as Bioproject data base under the accession number PRJNA604693.

Collection and processing of transcriptomic data

For the species *Nacella polaris* and *Cepaea nemoralis*, raw reads from the SRA database (NCBI) were imported to Geneious R10 (Biomatters Ltd., Auckland, New Zealand) to assemble transcriptomes. New MT sequences were identified by blasting already known MT sequences from close relatives against the new transcriptomes. For most other species, MT peptide sequences of the diverse gastropod families were identified using the blastn tool at the NCBI platform (https://blast.ncbi.nlm.nih.gov/Blast.cgi) against the database transcriptome shotgun assembly for gastropod species (taxid: 6448).

MT sequence confirmation via long distance (LD) PCR and quantitative Real-time PCR

Gene-specific primers (**Table S2A**) were designed from identified MT sequences derived from transcriptomic data (see above). For PCR, a 50µl approach was set up using the Advantage 2 PCR System (Clontech, Takara Bio Europe, Saint-Germain-en-Laye, France) (**TableS2B**). PCR products were

separated on a 1.5% agarose gel (Biozym, Hessisch Oldendorf, Germany) and gene specific bands were excised. DNA was purified applying the QIAquick™ Gel Extraction Kit (Qiagen, Hilden, Germany), and cleaned samples were sent to Microysnth AG (Balgach, Switzerland) for Sanger-sequencing. When necessary, subsequent cloning was performed with the TOPO® TA Cloning® Kit for sequencing (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA). Insert containing plasmids were purified using the QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany) and sent to Microysnth AG (Balgach, Switzerland) for Sanger sequencing. Primer design and sequence analysis were performed applying CLC Main workbench 6.9 (Quiagen, Aarhus, Denmark).

For CdMT quantification of *Cornu aspersum*, cDNA of the controls and Cd-exposed individuals were measured in triplicates using a 7500 Real Time PCR Analyzer with Power SYBR® Green detection (Applied Biosystems™ by ThermoFisher Scientific, USA). Details on primer design and concentrations as well as establishment of the calibration curve are described in [26]. Total RNA was used as a reference for transcriptional quantification (see [29]).

Phylogenetic analysis

Alignments of MT amino acid sequences were done with MUSCLE v3.8.31 [30], and manual corrections were applied if deemed appropriate. Alignment length was variable and species-specific, with protein lengths between 50 and 180 amino acids. All alignments applied for the present tree calculations are reported as FASTA alignments (see **Alignments S1-5**). Phylogenetic tree reconstructions were performed with RAxML v8.2.8 (maximum likelihood ML, [31]) and MrBayes v3.2.6 (Bayesian inference BI, [32]). For ML with the model PROTGAMMAIWAG, 1000 - 10,000 inferences were calculated, and 1,000 bootstrap replicates. For BI, 10 million generations were calculated with rates=invgamma and aamodelpr=mixed, average standard deviation of split frequencies.

In addition, phylogenetic trees were also computed with a maximum likelihood (ML) approach with 500 bootstrap replicates, using the freely accessible programme platform SeaView (version 4.7) of PRABI-Doua , using default settings. Overall topologies between BI and ML trees were very similar, and the trees with the lowest number of polytomies are shown.

Mitochondrial COI sequences were manually aligned and checked for correct amino acid translation; the ribosomal 18SrDNA sequences were aligned using the SINA Alignment tool v. 1.2.11, based on the SILVA database [33] (**Alignment S1**). In all phylogenetic reconstructions gaps were treated as missing data. Four partitions were defined in the concatenated data, one for each codon position of COI and one for 18SrDNA. ML analysis was performed using IQ-tree [34] allowing for model estimation in each partition; node supports were calculated using 1000 non-parametric UltraFast Bootstraps. For BI the best-fitting substitution models were obtained with Modeltest 3.7. [35]: GTR+I+G achieved the best AIC and BIC values in all four partitions. BI was performed with MrBayes v3.2.6 allowing for unlinked parameter

estimation in each partition. Five million generations were performed and 25% burnin was chosen to discard data prior to convergence of runs (standard deviation of split-frequencies below 0.01). The ML tree and the BI tree (data not shown) revealed the same topology.

Cloning and recombinant expression of *MT* genes from *Pomacea bridgesii* and *Lottia gigantea*

Full-length synthetic cDNAs for *PbMT1* and *LgMT1* genes were synthesized by Integrated DNA Technologies Company (Coralville, IA, USA) and by Synbiotech (Monmouth Junction NJ, USA), respectively. Both cDNAs were cloned into the *E. coli* pGEX-4T-1 expression vector (GE Healthcare) as described elsewhere [13] with minor modifications. Cloned *PbMT1* and *LgMT1* cDNAs were sequenced with the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) at the Scientific and Technological Centers of the University of Barcelona (ABIPRISM 310, Applied Biosystems).

For heterologous expression of GST-MT fusion proteins, 500 mL of LB medium with 100 μg mL^{-1} ampicillin were inoculated with protease-deficient *E. coli* BL21 cells previously transformed with the *PbMT1* pGEX-4T-1 or *LgMT1* pGEX-4T-1 recombinant plasmids. After overnight growth at 37 °C/250 rpm, the cultures were used to inoculate 5 L of fresh LB-100 μg mL^{-1} ampicillin medium. Gene expression was induced with 100 μM isopropyl-β-D-thiogalactopyranoside (IPTG) for 3 hours (h). After the first 30 minutes (min) of induction, cultures were supplemented with $ZnCl_2$ (300 μM), $CdCl_2$ (300 μM) or $CuSO_4$ (500 μM) in order to generate metal–MT complexes. Cells were harvested by centrifugation for 5 min at 9100 g (7700 rpm), and bacterial pellets were suspended in 125 mL of ice-cold PBS (1.4 M NaCl, 27 mM KCl, 101 mM Na_2HPO_4, 18 mM KH_2PO_4 and 0.5% v/v β-mercaptoethanol). Resuspended cells were sonicated (Sonifier Ultrasonic Cell Disruptor) 8 min at voltage 6 with pulses of 0.6 seconds, and then centrifuged for 40 min at 17200 g (12000 rpm) and 4° C.

Purification of recombinant metal-MT complexes

Protein extracts containing GST–PbMT1 or GST–LgMT1 fusion proteins were incubated with glutathione sepharose beads (GE Healthcare) for 1 h at room temperature with gentle rotation. GST–MT fusion proteins bound to the sepharose beads were washed with 30 mL of cold 1xPBS bubbled with argon to prevent oxidation. After three washes, GST–MT fusion proteins were digested with thrombin (GE Healthcare, 25 U L^{-1} of culture) overnight at 17 °C, thus enabling separation of the metal–MT complexes from the GST that remained bound to the sepharose matrix. The eluted metal–MT complexes were concentrated with a 3 kDa Centripep Low Concentrator (Amicon, Merck), and fractionated on a Superdex-75 FPLC column (GE Healthcare) equilibrated with 20 mM Tris–HCl, pH 7.0, and run at 0.8 mL min^{-1}. The protein-containing fractions, identified by their absorbance at 254 nm, were pooled and stored at -80 °C until use.

Analysis of recombinantly expressed metal-MT complexes

For determination of the molecular mass of the metal complex species in solution, the metal-MT complexes produced by recombinant expression were analyzed by electrospray ionization mass spectrometry (ESI-MS). For that purpose, a Micro Tof-Q Instrument (Bruker Daltonics Gmbh, Bremen, Germany) with a time-of-flight analyzer (ESI-TOF MS) was utilized, calibrated with ESI-L Low Concentration Tuning Mix (Agilent Technologies, Santa Clara, CA, USA), and interfaced with a Series 1100 HPLC pump (Agilent Technologies) equipped with an autosampler, both controlled by the Compass Software. The experimental conditions for analysis of Zn and Cd proteins were as follows. 10-20 µL of the sample were injected at 40 µL/min using the capillary-counter electrode voltage at 5.0 kV and the desolvation temperature in the 90-110 °C range. For Cu containing samples the conditions used were milder, applying the capillary-counter electrode voltage at 4.0 kV and the desolvation temperature at 80 °C. Spectra were collected throughout an m/z range from 800 to 2500. The liquid carrier was a 90:10 mixture of 15 mM ammonium acetate and acetonitrile, pH 7.0. All samples were injected in duplicates to ensure reproducibility.

NMR and metal titration

Fully cadmium-loaded forms of *Littorina littorea* and *Helix pomatia* MTs were produced by recombinant expression and uniformly ^{15}N-labelled in *E. coli* cells as described previously [20]. To demetallate the proteins their solutions were acidified in three buffer exchange steps, adding demetallation solutions (pH 2.0, 10 or 20 mM MES or TRIS, 10 mM TCEP) using Amicon Ultra 3K Centrifugal Filter Devices (EMD Millipore). All solutions were carefully purged with argon prior to use. Titrations were performed in 20 or 50 mM MES (pH 6.0), MES (pH 7.0) and Tris (pH 7.0) buffers with 10 mM TCEP yielding very similar results. Metallation was followed by recording [^{15}N,^{1}H]-HSQC spectra or best-type [^{15}N,^{1}H]-HSQC spectra [36]. Spectra were analyzed and peaks integrated applying the program CcpNmr v.2.4.2. [37].

Measurements of ^{15}N transverse relaxation rates (R2) were performed using a HSQC-type version of the Carr Purcell Meiboom Gill (CPMG) experiment [38]. 32 scans were performed per increment and T2 delays of 0, 17, 34, 51, 68, 102, 119, 204 and 305 ms were used. The relaxation delay in all these experiments was set to 2 s. Spectra were recorded using Zn_6- or Cd_6-HpMT and Zn_9- or Cd_9-LlMT samples. Zn-loaded MTs were generated by adding Zn to demetallated MTs. Peaks were integrated batchwise using the program SPSCAN and R2 rates extracted from least square fits to the standard exponential decay function with gnuplot.

All spectra were recorded at 298 K on a Bruker NEON 600 MHz or 700 MHz NMR spectrometer using a PRODIGY triple-resonance probe for samples at a concentration range of 0.1- 0.5 mM.

304 **Tissue sample digestion and metal analyses**

305 For Cd analysis, midgut gland tissue samples and lettuce leaves (*Lactuca sativa*) were oven-dried
306 at 60°C. After dry weight (d.w.) determination, samples were wet-digested at 70°C with a mixture of
307 HNO_3 (suprapur, Merck, Darmstadt, Germany) and deionized water (1:1) in 12 ml screw-capped
308 polyethylene tubes (Greiner, Kremsmünster, Austria). For complete oxidation, a few drops of H_2O_2 were
309 added to the hot digested samples. They were filled up with deionized water to a final volume of 11.5 ml.
310 Cd concentrations were measured by flame (Model 2380, Perkin Elmer, Boston, MA) or graphite furnace
311 atomic absorption spectrophotometry with polarized Zeeman background correction (Model Z-8200,
312 Hitachi, Japan) and $Pd(NO_3)_2$ as a matrix modifier, depending on concentration levels in the samples.
313 Calibration was performed with diluted titrisol standard solutions (Merck) prepared with de-ionized water
314 and 5% HNO_3 (suprapur, Merck). Lobster hepatopancreas powder (TORT-2, National Research Council,
315 Canada) was used as a standard reference material and processed in the same way as the samples (n = 5).
316
317 **Preparation and chromatography of *in vivo* MTs**

318 Purification and preparation of *in vivo* MTs for determination of molar metal ratios were performed
319 on centrifuged supernatants of midgut gland homogenates obtained from Cd-exposed snails (*Helix*
320 *pomatia*, *Cornu aspersum*) und slugs (*Arion vulgaris*), by applying successive fractionation steps on gel
321 permeation chromatography, anion exchange chromatography, ultrafiltration and Reverse phase HPLC [23].
322 For each species, HPLC fractions of the eluted MT peak were pooled and diluted 1:10 with deionized
323 water under addition of 1% HNO_3. Metal concentrations (Cd, Cu, Zn) were analysed in triplicate in 1 ml
324 aliquots by graphite furnace atomic absorption spectrophotometry with polarized Zeeman background
325 correction (Model Z-8200, Hitachi, Japan) as described above.
326
327 **Statistics**

328 Data from q-RT-PCR and metal analyses were evaluated using SigmaPlot 12.5 (SYSTAT
329 software, San Jose, CA, USA). Values were tested for normal distribution with the Shapiro–Wilk
330 normality test and the equal variance test. Outliers of normally distributed data were assessed with the
331 Grubbs test (https://www.graphpad.com/quickcalcs/Grubbs1.cfm). For not normally distributed data, non-
332 parametric methods (Mann-Whitney U test) were applied. Significance levels were set at $p \leq 0.05$.
333
334 **Results and Discussion**

335 In this work we propose that Cd acts as a driver in the evolution of gastropod metallothioneins. In
336 what follows we first describe the variety of gastropod MTs (section 1), structural features of MTs (section
337 2), and our phylogenetic analysis of how gastropod MTs changed during evolution to gain (section 3) or

loose (section 4) Cd-binding selectivity. Moreover, we describe how Cd-selectivity was achieved during evolution (section 5) and conclude with how changes in environmental Cd levels influenced Cd selectivity (section 6).

1. Gastropod diversity

Modern Gastropoda represent five distinct clades with about 80,000 species: Patellogastropoda, Neritimorpha, Vetigastropoda, Caenogastropoda and Heterobranchia. Their phylogenetic relationships [25,39–44] served as a reference for this study. Interestingly, several gastropod lineages have independently abandoned marine realms and successfully adapted to semi-terrestrial, terrestrial and freshwater environments [45]. This manifold colonization of non-marine habitats has promoted the huge diversity of gastropod traits, including the structural and functional diversity of their *MT* genes and proteins. Species and their MT sequences used for phylogenetic tree constructions of the present study are reported in **Table S3**.

2. Gastropod MTs: Structures, domain organization and metal binding features

Examples of primary MT structures across all major gastropod clades are shown in **Figure 1A-E**. Amino acid sequences of most gastropod MTs reflect a bipartite organization of one N-terminal metal-binding domain linked to a distinctly different C-terminal metal-binding domain **(Figure 1A, B, E)**. This kind of structural organization has been confirmed by NMR studies and molecular modeling [16,46]. It is therefore assumed that the primordial gastropod MT was a bidominial MT. Both N-terminal and C-terminal domains, contain nine Cys residues each which bind in a stoichiometric ratio, three divalent (mainly Cd^{2+}, Zn^{2+}), or six monovalent (mainly Cu^{+}) metal ions, such that a prototypical two-domain snail MT can accommodate six divalent or 12 monovalent metal ions, respectively [14]. An exception from this rule is observed in MTs of Patellogastropoda such as *Lottia gigantea* and *Patella vulgata*, which possess a deviating N-terminal domain that contains 10 instead of nine Cys residues, most of them arranged in form of double (Cys-Cys) motifs (**Figure 1B**). In addition to this, the N-terminal MT domain in some snail species has been duplicated once or several times independently, as seen in *Littorina littorea* and *Pomatias elegans* [16,20] **(Figure 1B)**. In the land snail *Alinda biplicata* and in some other species, tandem duplications generated multi-domain MTs (md-MTs) consisting of modular strings of up to nine N-terminal domain repeats, always linked to a single C-terminal domain that has, to the best of our knowledge, never been duplicated [47] **(Figure 1C)**. Domain duplications were also reported from bivalve MTs [48], which have probably emerged independently from those in gastropods. In gastropods, md-MTs can bind additional metal ions according to the number of added domains within the protein chain. For example, in In the three-domain MT of *Littorina littorea*, the metal binding ratio for Cd^{2+} has been extended to a number of nine Cd^{2+} ions as compared with six Cd^{2+} ions in normal bidominial snail MTs

[20,22]. Apart from this, N- and/or C-terminal domains are modified in some species by deletions at specific positions or through premature chain truncations **(Figure 1 D,E)**.

Across all gastropod MTs, primary structures of C-terminal domains appear to be higher conserved compared with N-terminal domains. A BLAST comparison of the N-terminal domain of *Littorina littorea* with those of all other gastropod CdMTs reveals low degrees of homology in a clade-specific gradation from Heterobranchia through Caenogastropoda to Patellogastropoda (**Figure 2**). In contrast, similarity scores between C-terminal domains of the same MT sequences are much more significant. This clearly demonstrates the higher degree of conservation of C-terminal against N-terminal metal binding domains, which is also confirmed by a distance matrix derived from single domain alignments (**Table S4**). We used *Littorina littorea* as a reference because this species possesses a well characterized CdMT [20,22] and occupies a central position between ancient and modern Gastropoda [41].

The higher evolutionary pressure for sequence conservation of the C-terminal domain in snail MTs is probably related to preferred Cd^{2+} loading into that part of the protein. This is demonstrated by NMR data of experiments, in which Cd^{2+} equivalents were added stepwise to the apo-MT of *Littorina littorea* **(Figure 3)**. Unlike the Cd-loaded MT **(Figure 3B)**, the apo-MT is unfolded and does not assume a specific three-dimensional shape (**Figure 3A**) [11,49]. Added Cd^{2+} is initially cooperatively incorporated into the C-terminal domain to build the C-terminal cluster (**Figure 3C**), before the two N-terminal domains form simultaneously (**Figure 3D**). This proves a clear priority for Cd^{2+} uptake into the C-terminal domain.

So far, the tertiary structure of two snail MTs has been disclosed by solution NMR, namely for the bidominial CdMT of the Roman snail, *Helix pomatia* [46], and for the three-domain CdMT of the periwinkle, *Littorina littorea* [20]. The tertiary structure of the Roman snail CdMT in its dumbbell shape reesembles the very similar structures of vertebrate MTs [50,51]. However, the metal-binding stoichiometry of the snail MT with six Cd^{2+} ions for the entire protein and three Cd^{2+} ions coordinated by nine Cys residues within each of the two domains, respectively, differs from the well-known metal binding stoichiometry of MTs from most other animal clades [46]. In mammalian MTs, four divalent metal ions are coordinated by 11 Cys residues in the C-terminal cluster (called alpha domain), whereas three divalent metal ions are bound by nine Cys residues in the N-terminal domain (called beta domain) [51]. The MT of *Littorina littorea* is the first reported animal MT ever that exhibits a three-domain partition [20].

Many snail MTs possess Cd- or Cu-selective binding preferences, and can be isolated as stable, homometallic metal complexes from native snail tissues [14,52]. Although the exact prerequisits for metal-selectivity are not yet fully understood, it appears that the frequency and position of certain non-coordinating amino acid residues in the primary sequence and their spatial arrangement in the tertiary structure are crucial determinants in conferring metal-selectivity to snail MTs [15,19–21].

The homometallic composition of metal-selective snail MTs was demonstrated by electrospray ionization mass spectrometry (ESI-MS) in recombinantly expressed and purified MT proteins [15,22].

Thereby, metal-selective MTs can be detected as homometallic complexes with their cognate metal ions (mainly Cd^{2+} or Cu^{+}) but appear as a heterometallic mixture of complexes with variable stoichiometry when forced to associate with non-cognate metal species [15,18]. In contrast, metal-unspecific snail MTs form heterometallic mixtures of complexes with variable stoichiometry in presence of any metal ions [17]. According to this definition, it appears that Cd-selective MTs have not equally evolved in all gastropod clades. For example, vetigastropod species like *Megathura crenulata* possess an unspecific MT that produces a mixture of sulfide containing heterometallic complexes with reduced stability when recombinantly expressed in Cd-enriched media [17] (**Figure 4**). In contrast, recombinant CdMTs of some Caenogastropoda (like *Littorina littorea*) and Heterobranchia (like *Helix pomatia*) form homometallic Cd^{2+} complexes (**Figure 5**). MTs of Patellogastropoda are also Cd-selective. However, because of the divergent primary structure of their N-terminal domain with 10 Cys residues (**Figure 1B**), CdMTs of Patellogastropoda bind seven instead of six Cd^{2+} ions per protein molecule, as demonstrated for *Lottia gigantea* (**Figure 5**). CdMTs of *Helix pomatia* and *Arion vulgaris* (Heterobranchia) bind six Cd^{2+} ions per protein molecule or nine Cd^{2+} ions in three-domain CdMTs like that of *Littorina littorea* (Caenogastropoda) (**Figure 5**). All gastropod CdMTs are incapable to form homometallic Cu^{+} complexes (**Figure 5**). However, due to the chemical similarity between Zn and Cd, some recombinant gastropod CdMTs can form homometallic complexes with divalent Zn^{2+} ions. This Zn^{2+}-binding selectivity is low in CdMTs of Patellogastropoda, as demonstrated for *Lottia gigantea* (**Figure 5**). In contrast, Zn preference is high for Caenogastropoda CdMTs (*Littorina littorea*) and Stylommatophora CdMTs (*Helix pomatia* and *Arion vulgaris*), which are able to form homometallic Zn^{2+} complexes in the presence of excessive Zn^{2+} concentrations, with the same stoichiometry as for Cd^{2+} (**Figure 5**). Interestingly, some evidence indicates Zn specificity in MTs of some mussels (Bivalvia), the mollusk sister class of gastropods [44].

3. Phylogeny suggests convergent evolution of CdMTs in early gastropod clades

The multitude of published and collected primary MT sequences from species across all clades of Gastropoda **(see Table S3)** and basic knowledge about their structure and metal-binding features (see above) fosters an attempt towards establishing a phylogeny of gastropod MTs and, in particular, Cd-selective snail MTs. The smallness of most MT proteins and the fact that the abundance of conserved cysteine residues and repeat motifs do not bear much phylogenetically evaluable information creates a challenge in such an analysis. In the present study, we have developed a domain and metal-specific approach to compensate somewhat for these handicaps.

Yet, confronting a phylogeny of neutral DNA markers with one based on Cd-selective MTs (**Figure 6**) reflects the evolution of Cd selectivity in MTs of three gastropod clades: Patellogastropoda, Caenogastropoda and Heterobranchia. It appears that Cd-selective MTs are predominantly observed in species that have adapted to littoral and terrestrial environments (**Figure 6**). A closer phylogenetic view

in which MTs of Panpulmonata (a taxon of Heterobranchia comprising the lineages of Sacoglossa, Syphonariodea, Hygrophila and Stylommatophora) [53] are rooted with MTs of Caneogastropoda (**Figure 7**) reveals the loss of Cd-selective MTs in freshwater snails of Caenogastropoda and Heterobranchia, and the initial emergence and subsequent loss of Cu-selective MT isoforms (CuMTs) in the lineage of Stylommatophora (terrestrial snails and slugs). In that context it is of interest that it was previously shown that snail CuMTs are involved in Cu regulation, possibly linked to hemocyanin synthesis [54,55].

Chronograms show that Cd selectivity developed from ancestral MTs twice independently. CdMTs evolved first in Patellogastropoda, about 430 million years (My) ago (**Figure 8A**). A second line of CdMTs emerged in the two sister clades Caenogastropoda and Heterobranchia, before 418 My ago (**Figure 8A)**. Apart from the phylogenetic evidence, another clear indication of this independent evolution is the emergence of a new kind of N-terminal metal-binding domain in CdMTs of Patellogastropoda (see above), which differs fundamentally from N-terminal domains in CdMTs of all other gastropod lineages **(Figure 1B)**. The two sister clades Caenogastropoda and Heterobranchia have shaped their CdMTs through parallel evolution (**Figure 8A, B**). This is reflected by sequence similarities and homologous domain organization across primary structures of their CdMTs **(Figure 1B)**.

Also indicated in the chronogram are some of the main mass extinction events through the evolutionary history of the earth (**Figure 8B**). Fluctuating emissions of Cd through continental and super-volcanic emissions in combination with these catastrophic extinction events [2,56–59] may have triggered convergent evolution of Cd-selective MTs in gastropod clades since 430 My ago (**Figure 8B**). Evidence for increased Cd emissions through geological eras is provided by elevated Cd concentrations in worldwide bedrock formations of different geological origin, from Paleozoic [60–63] through Mesozoic [63,64] and Cenozoic [65].

Based on experimental data with recombinant proteins [17], it appears that Cd selectivity is lacking in MTs of Vetigastropoda (**Figures 8A**), which forms a sister clade to Patellogastropoda [43]. The metal-specific character in MTs of Neritimorpha, on the other hand, is still unknown (**Figure 8A**). Since Neritimorpha from a sister clade to Caenogastropoda and Heterobranchia [43], it could be speculated that they share Cd-specific features with their two sister clades. On the other hand, a high degree of identity in primary sequence and domain organization between MTs of Vetigastropoda and Neritimorpha (**Figure 1A**) suggests the possibility that Neritimorpha MTs share some of their metal-binding features with those of Vetigastropoda. Future experiments through recombinant expression and ESI-MS analyses will probably resolve this question. The supposed zinc (Zn) specificity in MTs of some mussels (Bivalvia), the mollusk sister class of gastropods, is also indicated in **Figure 8A.** However, this evidence is scarce, being derived from one single experimental study [44].

4. Diversification and loss of cadmium selectivity during late gastropod radiation

Since the Cretaceous period, Cd selectivity of MTs was apparently lost in snail lineages that adapted to freshwater habitats. Accordingly, metal-binding features of MTs from *Pomatia bridgesii* (family of Ampullariidae, Caenogastropoda) and *Biomphalaria glabrata* (Hygrophila, Heterobranchia) resemble those of the unspecific *Megathura crenulata* MT [17] (**Figure 4**). As indicated by their primary sequence and domain organization (**Figure 1D, E**), a loss of Cd selectivity may also have occurred in caenogastropod species of freshwater Calyptreidae and Buccinidae (**Figure 6**). The loss of Cd selectivity in these MTs is a derived character (**Figure 7**), suggesting that metal selectivity was no longer required in MTs of freshwater snails. In some freshwater species of Caenogastropoda such as *Pomacea canaliculata*, MTs have developed N-terminal repeats, similar to some snail CdMTs (**Figure 1C**).

In terrestrial snails of Stylommatophora (Heterobranchia), gene duplications of the primordial CdMT led to the emergence of three MT isoforms, each of them devoted to different, metal-specific tasks [14,55,66]. First, a gene duplication of *CdMT* gave rise to Cu-selective MTs, which form homometallic Cu^{+} complexes at a ratio of 12 Cu^{+} ions per protein molecule, but neither bind Cd^{2+} nor Zn^{2+} (**Figure S1, S2**). In a second event of gene duplication, *CuMT* genes lost their Cu selectivity in the so-called CdCuMT isoforms [18,19,55,66–68] (**Figure S1 S2**). Phylogenetic trees (**Figure S3**) support the chronological succession of these evolutionary steps.

5. Evolutionary optimization of Cd selectivity and specificity

For the sake of clarity, we like to distinguish between Cd (or metal) selectivity and specificity of MTs. We define Cd *selectivity* as the binding preference of an MT for Cd^{2+} ions in presence of other metal ions, mainly Zn^{2+} and Cu^{+}. We define Cd (or metal) *specificity* as the involvement of the respective MT into a Cd- or metal-specific physiological function, which is often the consequence of its metal binding selectivity. For example, Cd-selective snail MTs are predominantly involved in Cd-specific functions like detoxification [14,69,70].

Accordingly, we can observe that during gastropod evolution both, metal selectivity and physiological specificity of snail CdMTs have been optimized in favor of Cd^{2+}. The CdMT of *Littorina littorea*, for example, has been optimized for Cd^{2+} complexation to the disadvantage of Zn^{2+} binding. This can be concluded indirectly from the better fit of the protein backbone to the Cd vs the Zn cluster. To this end we measured ^{15}N dynamics NMR data that probe for rigidity of the polypeptide backbone. Transverse relaxation (R2) rates of Zn^{2+}-loaded CdMTs are increased by 14 and 8 Hz in the N-terminal N1 and N2 domains of the CdMT of *Littorina littorea,* respectively, and by up to 5 Hz in the C-terminal domain of the *Helix pomatia* CdMT (**Figure 9**) when compared to the Cd^{2+}-loaded forms. The increase in transverse relaxation rates reflects additional contributions from conformational exchange only for the Zn^{2+} species, indicating that the complexes with the cognate Cd^{2+} ion are conformationally more stable (**Figure 9**). Similarly, NMR studies of the CdMT of *Helix pomatia* indicate a structural optimization for Cd^{2+} rather

than Zn^{2+} binding [46]. Further evidence for an evolutionary optimization for Cd binding in these MTs comes from the fact that Cd^{2+} ions are incorporated into these proteins cooperatively (**Figure 3**). The fact that peaks in the [^{15}N,^{1}H]-HSQC spectra occur at the same positions as in the fully-metallated domains indicates that no partially metallated domains form in situations of substoichiometric metal content. Strikingly, the Cd-selective character of gastropod CdMTs is also maintained in the presence of Cu^{+} ions at equimolar concentrations with Cd^{2+}, as demonstrated recently for the recombinantly produced CdMT (AvMT1) of the terrestrial slug, *Arion vulgaris* [23]. This is remarkable since the evolution of thiolate proteins with an apparent preference for binding Cd^{2+} over Cu^{+} is a particular feature of snail CdMTs which is otherwise not observed in other animal MTs [46,50,51], and seems to contradict the chemical rules of the Irving Williams series [71]. These rules predict that the stability constants of transition-metal ion complexes increase by a factor of 100 to 1000 from Cd- towards Cu-thiolates [72,73]. However, The Irving Williams rules may not apply to metal-selective snail MTs, considering that they do not contain simple binary metal-thiolate complexes. In the CdMTs of *Littorina littorea* and *Helix pomatia*, for example, divalent Cd^{2+} ions are tetrahedrally coordinated [20,46], forming Cd-thiolate clusters that most likely differ in their structural configuration from the Cu-thiolate clusters of snail CuMTs [14]. Importantly, it was demonstrated that the replacement of a few amino acid positions in the near vicinity to the metal-coordinating Cys residues can have a strong impact on the metal binding preferences of snail MTs [15,19,21], probably due to spatial and charge-dependent constraints upon formation of protein-metal complexes. We suspect that such amino acid replacements must have gradually improved/modified the Cd-binding selectivity of snail MTs during evolution. Apart from this, the capacity for Cd-loading of many snail CdMTs has been increased through evolutionary multiplication of Cd-binding domains as demonstrated for the littoral periwinkle, *Littorina littorea*, and the land snails *Pomatias elegans* and *Alinda biplicata* [16,20,47]. At the functional level, evolutionary optimization for Cd binding in CdMTs has resulted in Cd-specific detoxification pathways within snail tissues. This is reflected by the fact that native purified gastropod CdMTs contain mainly Cd^{2+}, but only small amounts of Zn^{2+} or Cu^{+} (**Figure 10A**). Concomitantly, Cd inactivation in these species is enhanced by metal-dependent upregulation of the respective *CdMT* genes (**Figure 10B**), and tissue or cell-specific expression of *CdMT* mRNA [23,69,74].

6. Environmental Cd levels through the earth history: an important evolutionary driver

Cd is carcinogenic and highly toxic in animals, even at low concentrations [75]. The chemical similarity of this metal and its frequent co-occurrence with Zn in ore deposits of the earth crust make Cd a dangerous competitor for Zn-dependent cellular processes [5]. Cd can also compete with calcium (Ca) [76], and hence affect gastropods that depend on Ca pathways for bio-mineralization of their shells [77].

Gain of Cd-selective MTs may have provided an advantage particularly for gastropod lineages that have adapted to littoral, semi-terrestrial, and terrestrial conditions. Recent natural Cd concentrations in

seawater follow those of algal nutrients such as phosphate [3], displaying higher concentrations in deeper oceanic waters and exhibit a depletion towards neritic surface waters down to concentrations of 10^{-9} – 10^{-10} M [78,79] (**Figure 11**). Complex formation of Cd^{2+} with inorganic and organic ligands further decreases its biological availability in neritic seawater realms [80]. The situation changes drastically in the littoral zone, where marine habitats come into contact with the continental earth crust, in which natural Cd background concentrations, at 10^{-8} – 10^{-7} M, can be up to 100 times higher than those of superficial seawater [81] (**Figure 11**). Decreasing seawater salinities in the supra-littoral zone can even enhance the availability of Cd^{2+} for animals [82,83].

Gastropods of these habitats have adapted to fluctuating environmental conditions [84] but also had to cope with increasing Cd concentrations. Inactivation of Cd^{2+} ions by metal-selective MTs would, therefore, confer on them a physiological advantage [85] over energy-consuming activities for continuously re-adjusting intracellular Cd concentrations [86]. Upon adaptation to terrestrial life, gastropods have learned to cope with alternating and adverse environmental conditions [87,88]. Hence, the conservation of Cd-selective MTs may also be beneficial for land snails [16] (**Figure 11**).

In contrast to terrestrial snails, freshwater species of Caenogastropoda and Heterobranchia have lost their Cd binding selectivity, likely because natural Cd background concentrations in freshwater habitats with about 10^{-10} – 10^{-12} M are the lowest of any snail habitat on earth [89] (**Figure 11**).

The multitude of metal-selective MT variants naturally occurring in snails offers the unique possibility to apply them as models for optimization of MT metal binding features through experiments in the laboratory. This may promote our understanding of about how amino acid replacements modify metal selectivity in MTs, and could have implications for the design of novel artificial Cd-binding proteins for the sake of basic research or for application in environmental bioremediation [90,91]. It underscores once more the true model character of metal-selective snail MTs.

7. Conclusions

Some important conclusions are derived from our findings: First, presence of Cd has been a continuous evolutionary stimulus through the last 430 million years, driving convergent evolution and optimization of Cd-selective MTs in gastropod clades. Second, the C-terminal domain of Cd-selective gastropod MTs has been subjected to a high pressure for evolutionary conservation, which we attribute to its important role for immobilizing Cd^{2+}. Third, gastropod adaptation to habitats with different Cd background levels has triggered MT diversification towards partial or complete loss of metal selectivity. Fourth, the natural evolution in snails of an array of differently metal-selective MT variants designates them MTs as model molecules and indicates that it is possible to design artificial Cd-selective peptides.

Acknowledgements

583 This work was funded by a cooperation grant to R.D. and O.Z. from the Austrian Science Fund
584 and the Swiss National Science Foundation (DACH grant No I 1482-N28), and a grant from the
585 Forschungskredit of the University of Zurich to C.B. (grant no. FK-18-083). B.E. was supported by a grant
586 for young scientists at the University of Innsbruck. Researchers from Barcelona want to acknowledge the
587 Spanish Ministerio de Ciencia e Innovación and FEDER for the projects BIO2015-67358-C2-2-P (M.C.
588 and O.P.) and BIO2015-67358-C2-1-P (R.A.). M.C. and O.P. are members of the "Grup de Recerca de la
589 Generalitat de Catalunya", ref. 2017SGR-864.
590 We thank Thomas Ostermann for help with the transcriptome assembly and Philipp Andesner
591 (both: University of Innsbruck, Austria) for producing the neutral marker sequences. Transcriptomic data
592 have been achieved in part using the HPC infrastructure of the University of Innsbruck. We also thank the
593 Servei d'Anàlisi Química (SAQ) at the Universitat Autònoma de Barcelona (ICP-AES, CD, UV-vis, ESI-
594 MS) for allocating instrument time and three anonymous reviewers for constructive comments on an
595 earlier version of the manuscript.
596
597

References

1 K. H. Wedepohl, *Geochim. Cosmochim. Acta*, 1995, **59**, 1217–1232.

2 J. T. Cullen and M. T. Maldonado, in *Cadmium: From Toxicity to Essentiality. Metal Ions in Life Sciences*, eds. A. Sigel, H. Sigel and R. Sigel, Springer, Dordrecht, 2013, pp. 31–62.

3 T. W. Lane and F. M. M. Morel, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 4627–4631.

4 T. W. Lane, M. A. Saito, G. N. George, I. J. Pickering, R. C. Prince and F. M. M. Morel, *Nature*, 2005, **434**, 42.

5 A. Martelli, E. Rousselet, C. Dycke, A. Bouron and J. M. Moulis, *Biochimie*, 2006, **88**, 1807–1814.

6 A. M. Sandbichler and M. Höckner, *Int. J. Mol. Sci.*, , DOI:10.3390/ijms17010139.

7 C. D. Klaassen, J. Liu and S. Choudhuri, *Annu. Rev. Pharmacol. Toxicol.*, 1999, **39**, 267–294.

8 R. D. Palmiter, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 8428–8430.

9 S. G. Bell and B. L. Vallee, *ChemBioChem*, 2009, **10**, 55–62.

10 A. Krężel and W. Maret, *Int. J. Mol. Sci.*, 2017, **18**, 1237.

11 C. A. Blindauer and O. I. Leszczyszyn, *Nat. Prod. Rep.*, 2010, **27**, 720–741.

12 G. Isani and E. Carpenè, *Biomolecules*, 2014, **4**, 435–457.

13 S. Calatayud, M. Garcia-Risco, N. S. Rojas, L. Espinosa-Sánchez, S. Artime, O. Palacios, C. Cañestro and R. Albalat, *Metallomics*, 2018, **10**, 1585–1594.

14 R. Dallinger, B. Berger, P. Hunziker and J. H. R. Kägi, *Nature*, 1997, **388**, 237–238.

15 Ò. Palacios, A. Pagani, S. Pérez-Rafael, M. Egg, M. Höckner, A. Brandstätter, M. Capdevila, S. Atrian and R. Dallinger, *BMC Biol.*, , DOI:10.1186/1741-7007-9-4.

16 L. Schmielau, M. Dvorak, M. Niederwanger, N. Dobieszewski, V. Pedrini-Martha, P. Ladurner, J. Rodríguez-Guerra Pedregal, J. D. Maréchal and R. Dallinger, *Sci. Total Environ.*, 2019, **648**, 561–571.

17 S. Pérez-Rafael, A. Mezger, B. Lieb, R. Dallinger, M. Capdevila, Ò. Palacios and S. Atrian, *J. Inorg. Biochem.*, 2012, **108**, 84–90.

18 Ò. Palacios, S. Pérez-Rafael, A. Pagani, R. Dallinger, S. Atrian and M. Capdevila, *J. Biol. Inorg. Chem.*, 2014, **19**, 923–935.

19 S. Pérez-Rafael, F. Monteiro, R. Dallinger, S. Atrian, Ò. Palacios and M. Capdevila, *Biochim. Biophys. Acta - Proteins Proteomics*, 2014, **1844**, 1694–1707.

20 C. Baumann, A. Beil, S. Jurt, M. Niederwanger, O. Palacios, M. Capdevila, S. Atrian, R. Dallinger and O. Zerbe, *Angew. Chemie - Int. Ed.*, 2017, **56**, 4617–4622.

21 M. Niederwanger, S. Calatayud, O. Zerbe, S. Atrian, R. Albalat, M. Capdevila, Ò. Palacios and R. Dallinger, *Int. J. Mol. Sci.*, , DOI:10.3390/ijms18071457.

22 Ò. Palacios, E. Jiménez-Martí, M. Niederwanger, S. Gil-Moreno, O. Zerbe, S. Atrian, R. Dallinger and M. Capdevila, *Int. J. Mol. Sci.*, 2017, **18**, 1–16.

23 M. Dvorak, R. Lackner, M. Niederwanger, C. Rotondo, R. Schnegg, P. Ladurner, V. Pedrini-Martha, W. Salvenmoser, L. Kremser, H. Lindner, M. Garcia-Risco, S. Calatayud, R. Albalat, Ò. Palacios, M. Capdevila and R. Dallinger, *Metallomics*, 2018, **10**, 1638–1654.

24 R. Dallinger, M. Chabicovsky and B. Berger, *Environ. Toxicol. Chem.*, 2004, **23**, 890–901.

25 K. M. Jörger, I. Stöger, Y. Kano, H. Fukuda, T. Knebelsberger and M. Schrödl, *BMC Evol. Biol.*, , DOI:10.1186/1471-2148-10-323.

26 M. Höckner, K. Stefanon, D. Schuler, R. Fantur, A. De Vaufleury and R. Dallinger, *J. Exp. Zool. Part A Ecol. Genet. Physiol.*, 2009, **311**, 776–787.

27 M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. Di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman and A. Regev, *Nat. Biotechnol.*, 2011, **29**, 644–652.

28 A. Priyam, B. J. Woodcroft, V. Rai, A. Munagala, I. Moghul, F. Ter, M. A. Gibbins, H. Moon, G. Leonard, W. Rumpf and Y. Wurm, *BioRxiv*, 2015, 033142.

29 V. Pedrini-Martha, M. Niederwanger, R. Kopp, R. Schnegg and R. Dallinger, *PLoS One*, 2016, **11**, 1–19.

30 R. C. Edgar, *Nucleic Acids Res.*, 2004, **32**, 1792–7.

31 A. Stamatakis, *Bioinformatics*, 2014, **30**, 1312–1313.

32 F. Ronquist, M. Teslenko, P. Van Der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard and J. P. Huelsenbeck, *Syst. Biol.*, 2012, **61**, 539–542.

33 E. Pruesse, J. Peplies and F. O. Glöckner, *Bioinformatics*, 2012, **28**, 1823–1829.

34 J. Trifinopoulos, L. T. Nguyen, A. von Haeseler and B. Q. Minh, *Nucleic Acids Res.*, 2016, **44**, W232–W235.

35 D. Posada and K. A. Crandall, *Bioinformatics*, 1998, **14**, 817–818.

36 P. Schanda, E. Kupce and B. Brutscher, *J. Biomol. NMR*, 2005, **33**, 199–211.

37 W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides and E. D. Laue, *Proteins Struct. Funct. Bioinforma.*, 2005, **59**, 687–696.

38 S. Meiboom and D. Gill, *Rev. Sci. Instrum.*, 1958, **29**, 688–691.

39 A. Dinapoli and A. Klussmann-Kolb, *Mol. Phylogenet. Evol.*, 2010, **55**, 60–76.

40 B. Dayrat, M. Conrad, S. Balayan, T. R. White, C. Albrecht, R. Golding, S. R. Gomes, M. G. Harasewych and A. M. de Frias Martins, *Mol. Phylogenet. Evol.*, 2011, **59**, 425–437.

41 F. Zapata, N. G. Wilson, M. Howison, S. C. S. Andrade, K. M. Jörger, M. Schrödl, F. E. Goetz, G. Giribet and C. W. Dunn, *Proc. R. Soc. B*, 2014, **281**, 20141739.

42 O. Razkin, B. J. Gómez-Moliner, C. E. Prieto, A. Martínez-Ortí, J. R. Arrébola, B. Muñoz, L. J. Chueca and M. J. Madeira, *Mol. Phylogenet. Evol.*, 2015, **83**, 99–117.

43 T. J. Cunha and G. Giribet, *Proc. R. Soc. B Biol. Sci.*, 2019, **286**, 20182776.

44 H. Yang, J. Zhang, J. Xia, J. Yang, J. Guo, Z. Deng and M. Luo, *Int. J. Mol. Sci.*, 2018, **19**, 3646.

45 A. Klussmann-Kolb, A. Dinapoli, K. Kuhn, B. Streit and C. Albrecht, *BMC Evol. Biol.*, , DOI:10.1186/1471-2148-8-57.

46 A. Beil, S. Jurt, R. Walser, T. Schönhut, P. Güntert, Ò. Palacios, S. Atrian, M. Capdevila, R. Dallinger and O. Zerbe, *Biochemistry*, 2019, **58**, 4570–4581.

47 V. Pedrini-Martha, S. Köll, M. Dvorak and R. Dallinger, *Int. J. Mol. Sci.*, 2020, submitted.

48 R. Orihuela, J. Domènech, R. Bofill, C. You, E. A. Mackay, J. H. R. Kägi, M. Capdevila and S. Atrian, *J. Biol. Inorg. Chem.*, 2008, **13**, 801–812.

49 N. Romero-Isart and M. Vašák, *J. Inorg. Biochem.*, 2002, **88**, 388–396.

50 Kägi J.H.R., in *Metallothionein III: Biological Roles and Medical Implications, Proceedings of the Metallothionein International Conference, 1992, Tsukuba, Japan*, Birkhauser Verlag, Boston, 1993, vol. 3, pp. 29–56.

51 E. H. Fischer and E. W. Davie, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 3333–3334.

52 P. M. Gehrig, C. You, R. Dallinger, C. Gruber, M. Brouwer, J. H. R. Kägi and P. E. Hunziker, *Protein Sci.*, 2000, **9**, 395–402.

53 K. M. Kocot, K. M. Halanych and P. J. Krug, *Mol. Phylogenet. Evol.*, 2013, **69**, 764–771.

54 R. Dallinger, M. Chabicovsky, E. Hödl, C. Prem, P. Hunziker and C. Manzl, *Am. J. Physiol. Integr. Comp. Physiol.*, 2005, **289**, R1185–R1195.

55 M. Höckner, K. Stefanon, A. De Vaufleury, F. Monteiro, S. Pérez-Rafael, Ò. Palacios, M. Capdevila, S. Atrian and R. Dallinger, *Biometals*, 2011, **24**, 1079–1092.

56 E. B. Naimark and A. V. Markov, *Biol. Bull. Rev.*, 2011, **1**, 71–81.

57 S. E. Grasby, B. Beauchamp, D. P. G. Bond, P. Wignall, C. Talavera, J. M. Galloway, K. Piepjohn, L. Reinhardt and D. Blomeier, *Bull. Geol. Soc. Am.*, 2015, **127**, 1331–1347.

58 D. P. G. Bond and S. E. Grasby, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 2017, **478**, 3–29.

59 G. Keller, P. Mateo, J. Punekar, H. Khozyem, B. Gertsch, J. Spangenberg, A. M. Bitchong and T. Adatte, *Gondwana Res.*, 2018, **56**, 69–89.

60 S. Aizawa and H. Akaiwa, *Chem. Geol.*, 1992, **98**, 103–110.

61 E. Porębska and Z. Sawłowicz, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 1997, **132**, 343–354.

62 M. L. W. Tuttle, G. N. Breit and M. B. Goldhaber, *Appl. Geochemistry*, 2009, **24**, 1565–1578.

63 Y. Liu, T. Xiao, R. B. Perkins, J. Zhu, Z. Zhu, Y. Xiong and Z. Ning, *J. Geochemical Explor.*, 2017, **176**, 42–49.

64 H. J. Brumsack, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 2006, **232**, 344–361.

65 F. Batifol, C. Boutron and M. De Angelis, *Nature*, 1989, **337**, 544–546.

66 V. Pedrini-Martha, R. Schnegg, P. E. Baurand, A. deVaufleury and R. Dallinger, *Comp. Biochem. Physiol. Part - C Toxicol. Pharmacol.*, 2017, **199**, 38–47.

67 P. E. Baurand, V. Pedrini-Martha, A. De Vaufleury, M. Niederwanger, N. Capelli, R. Scheifler and R. Dallinger, *PLoS One*, 2015, **10**, 1–14.

68 P. E. Baurand, R. Dallinger, M. Niederwanger, N. Capelli, V. Pedrini-Martha and A. de Vaufleury, *Environ. Sci. Pollut. Res.*, 2016, **23**, 3062–3067.

69 M. Chabicovsky, H. Niederstätter, R. Thaler, E. Hödl, W. Parson, W. Rossmanith and R. Dallinger, *Toxicol. Appl. Pharmacol.*, 2003, **190**, 25–36.

70 F. Hispard, D. Schuler, A. De Vaufleury, R. Scheifler, P. M. Badot and R. Dallinger, *Environ. Toxicol. Chem.*, 2008, **27**, 1533–1542.

71 H. Irving and R. J. P. Williams, *J. Chem. Soc.*, 1953, 3192–3210.

72 M. Erk and B. Raspor, *Anal. Chim. Acta*, 1998, **360**, 189–194.

73 G. Meloni, V. Sonois, T. Delaine, L. Guilloreau, A. Gillet, J. Teissie, P. Faller and M. Vasak, *Nat. Chem. Biol.*, 2008, **4**, 366–372.

74 E. Hödl, E. Felder, M. Chabicovsky and R. Dallinger, *Cell Tissue Res.*, 2010, **341**, 159–171.

75 W. Maret and J. M. Moulis, in *Cadmium: From Toxicity to Essentiality. Metal Ions in Life Sciences*, eds. A. Sigel, H. Sigel and R. Sigel, Springer, Dordrecht, 2013, pp. 1–29.

76 G. Choong, Y. Liu and D. M. Templeton, *Chem. Biol. Interact.*, 2014, **211**, 54–65.

77 D. Faubel, M. Lopes-Lima, S. Freitas, L. Pereira, J. Andrade, A. Checa, H. Frank, T. Matsuda and J. Machado, *Mar. Freshw. Behav. Physiol.*, 2008, **41**, 93–108.

78 E. A. Boyle, F. Sclater and J. M. Edmond, *Nature*, 1976, **263**, 42–44.

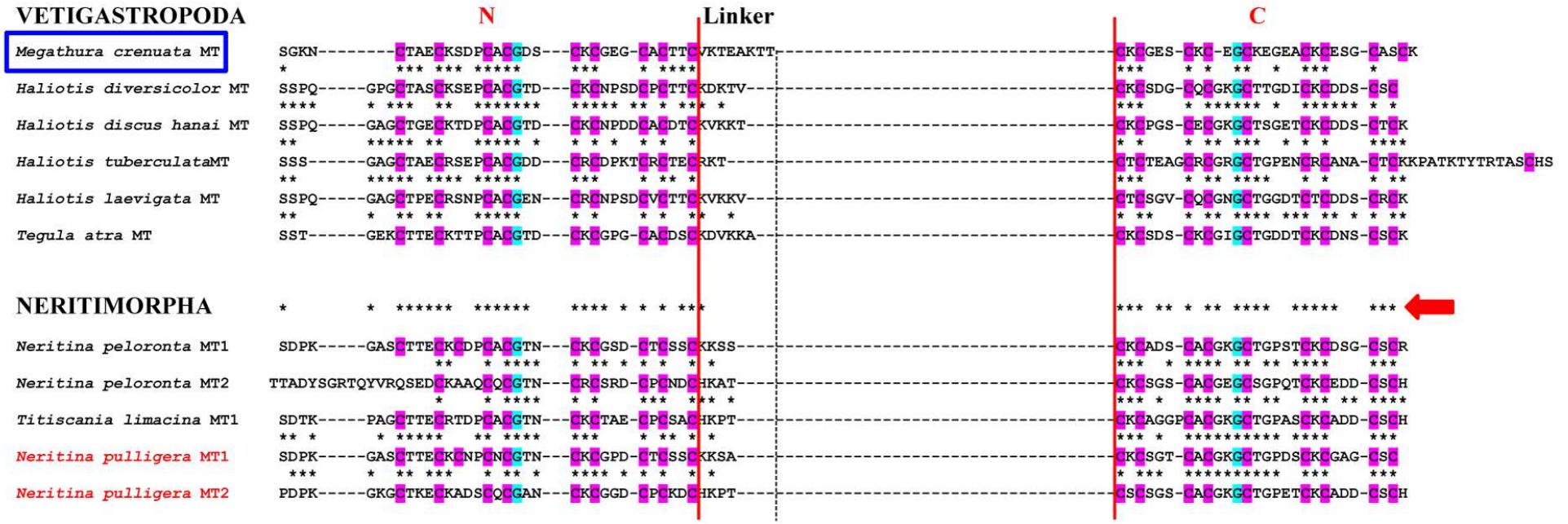
79 W. Abouchami, S. J. G. Galer, H. J. W. De Baar, R. Middag, D. Vance, Y. Zhao, M. Klunder, K. Mezger, H. Feldmann and M. O. Andreae, *Geochim. Cosmochim. Acta*, 2014, **127**, 348–367.

80 K. W. Bruland, *Limnol. Oceanogr.*, 1992, **37**, 1008–1017.

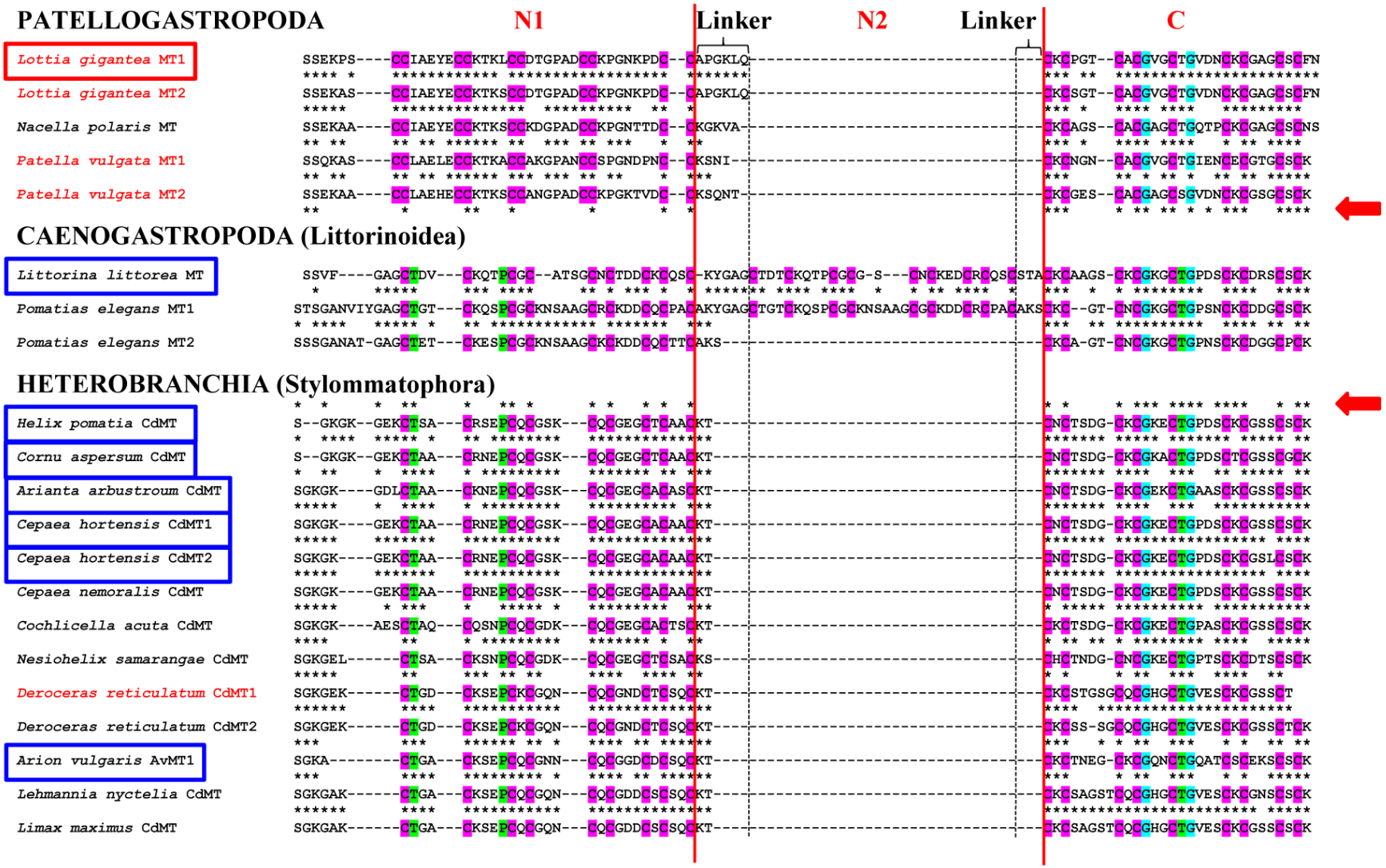
81 S. Sposito, in *Metal Ioins in Biological Systems, Vol. 20, Concepts on Metal ion Toxicity*, eds. H. Sigel and A. Sigel, Marcel Dekker Inc., New York, 1986, pp. 1–20.

82 P. Bjerregaard and M. H. Depledge, *Mar. Biol.*, 1994, **119**, 385–395.
83 G. Roesijadi, *Mar. Environ. Res.*, 1994, **38**, 147–168.
84 K. B. Storey, B. Lant, O. O. Anozie and J. M. Storey, *Comp. Biochem. Physiol. Part A*, 2013, **165**, 448–459.
85 T. E. English and K. B. Storey, *J. Exp. Biol.*, 2003, **206**, 2517–2524.
86 M. H. Depledge and P. S. Rainbow, *Comp. Biochem. Physiol.*, 1990, **97C**, 1–7.
87 S. Slotsbo, K. V. Fisker, L. M. Hansen and M. Holmstrup, *J. Comp. Physiol. B*, 2011, **181**, 1001–1009.
88 T. Bishop and M. D. Brand, *J. Exp. Biol.*, 2000, **203**, 3603–12.
89 C. A. Mebane, in *US Department of the Interior & US Geological Survey*, Scientific Investigations Report 2006-5245, Reston, Virginia, v. 1.2., 2010, p. 130.
90 A. F. A. Peacock and V. L. Pecoraro, in *Cadmium: From Toxicity to Essentiality (Metal Ions in Life Sciences Vol. 11)*, eds. A. Sigel, H. Sigel and R. Sigel, Springer, Dordrecht, 2013, pp. 303–338.
91 M. Mejáre and L. Bülow, *Trends Biotechnol.*, 2001, **19**, 67–73.
92 K. M. Kocot, J. T. Cannon, C. Todt, M. R. Citarella, A. B. Kohn, A. Meyer, S. R. Santos, C. Schander, L. L. Moroz, B. Lieb and K. M. Halanych, *Nature*, 2011, **477**, 452–456.
93 D. J. Colgan, W. F. Ponder, E. Beacham and J. Macaranas, *Mol. Phylogenet. Evol.*, 2007, **42**, 717–737.
94 C. M. Wade, P. B. Mordan and B. Clarke, *Proc. R. Soc. B*, 2001, **268**, 413–422.
95 E. V. Soldatenko and A. A. Petrov, *J. Morphol.*, 2019, **280**, 508–525.
96 M. Niederwanger, M. Dvorak, R. Schnegg, V. Pedrini-Martha, K. Bacher, M. Bidoli and R. Dallinger, *Int. J. Mol. Sci.*, 2017, **18**, 1747.
97 D. Benito, M. Niederwanger, U. Izagirre, R. Dallinger and M. Soto, *Int. J. Mol. Sci.*, 2017, **18**, 1815.

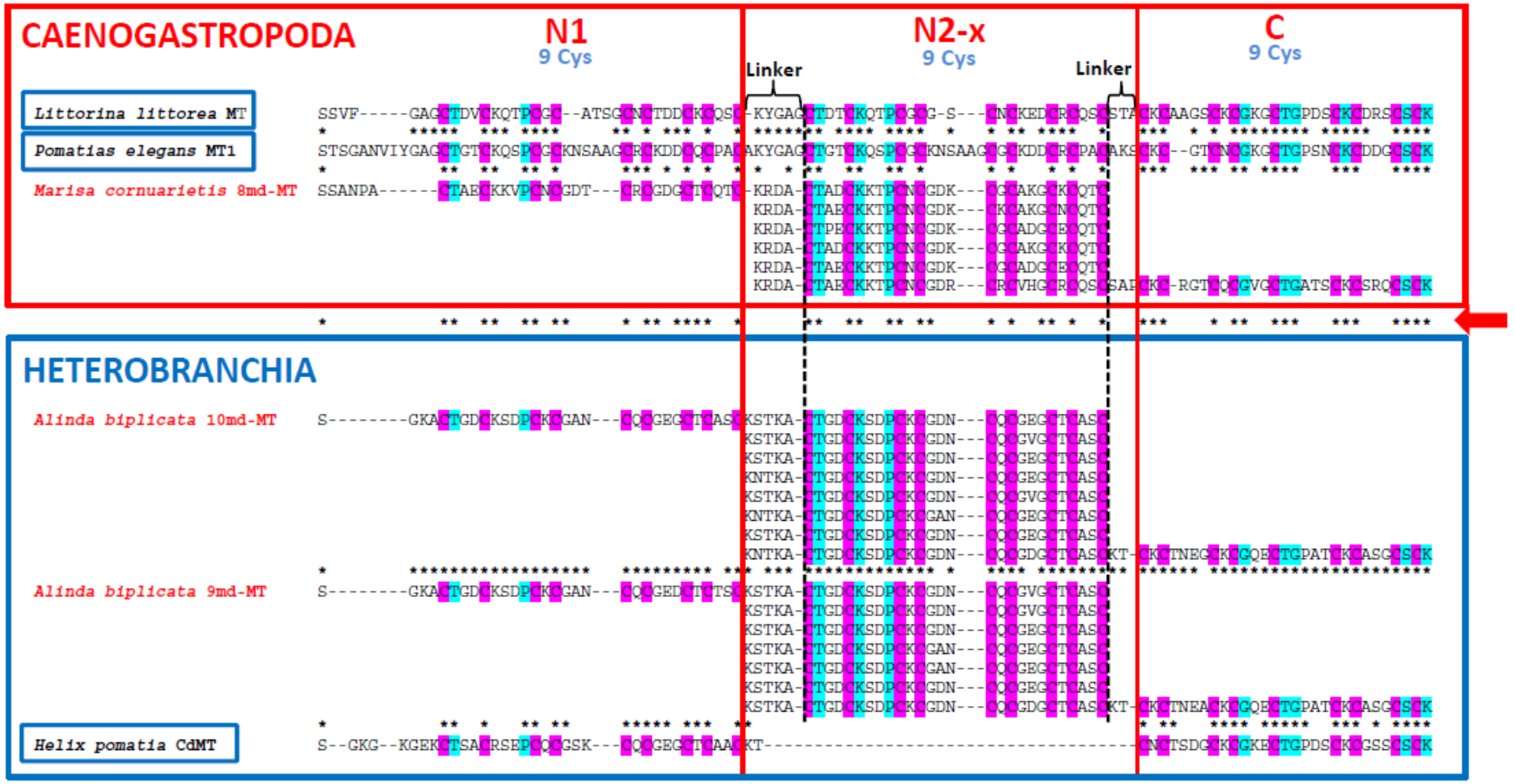
Figure 1A-E: Sequence alignments of snail metalothioneins.
Cys positions are underlaid in pink, conserved non-Cys positions through sequences of both clades are underlaid in light blue. Identical amino acid positions between pairwise aligned sequences are indicated by black stars. Domain boundaries of the N-terminal and the C-terminal domain (designated above the alignments with N and C) are indicatd by bold red lines. The linker between the two domains is shown in black letters, its boundary is symbolized by a dotted line. The gaps between the two domains were inserted indicating the lack of a second N-terminal domain (present in other gastropod MTs). MTs of species shown in red letters were sequenced in this study for the first time while sequences in black letters were downloaded from publications or databases. Species for which metal selectivity features of respective MTs were documented experimentally by us through MS or NMR methods elsewhere are framed in blue.



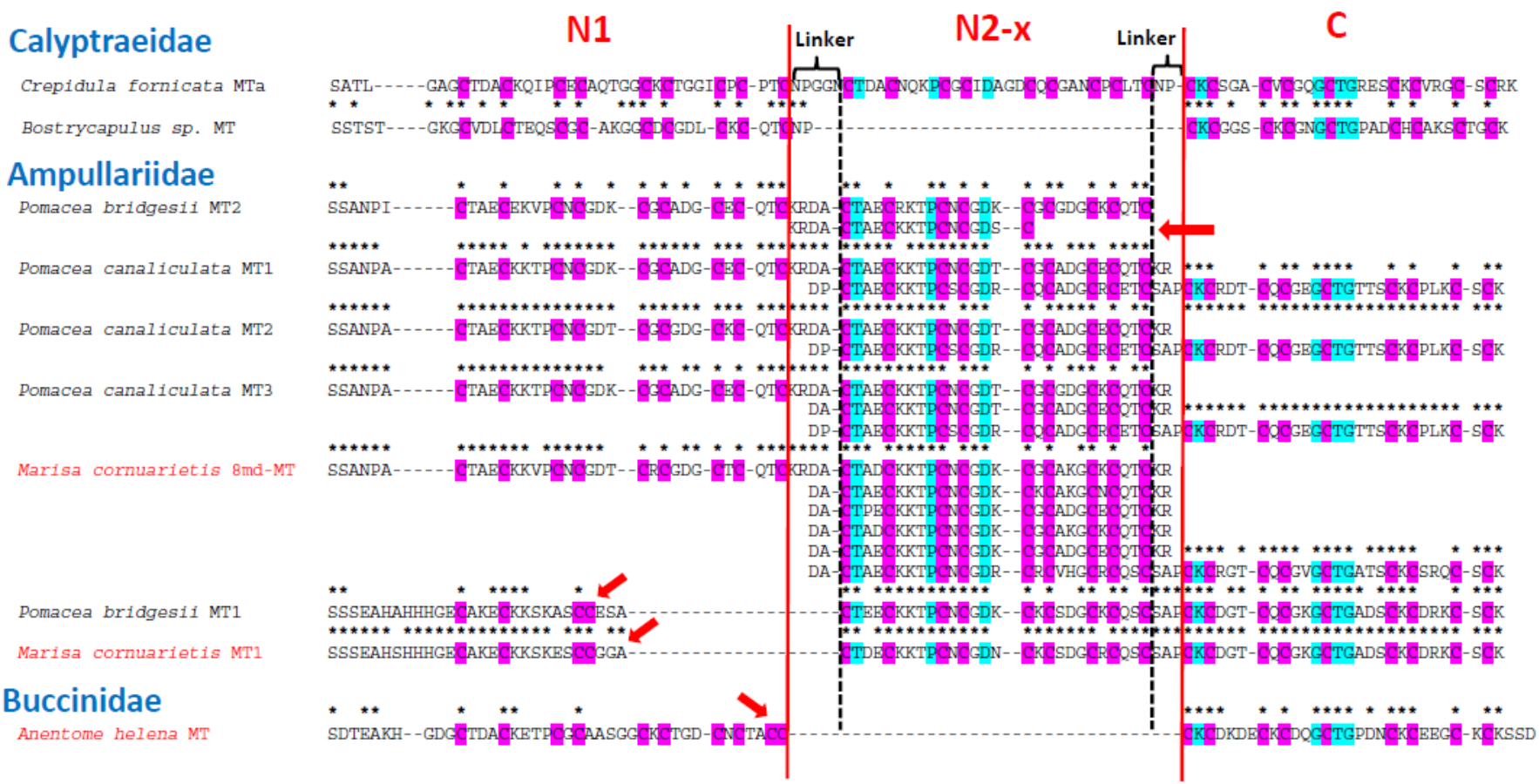
A Metal-binding domain organization and amino acid sequence alignment of unspecific MTs from the gastropod clade of Vetigastropoda and MTs with still unknown metal binding features from Neritimorpha. The bold red arrow on the right hand of the alignments points to sequence identities between MTs of Tegula atra (Vetigastropoda) and Neritina peloronta MT1 (Neritimorpha).



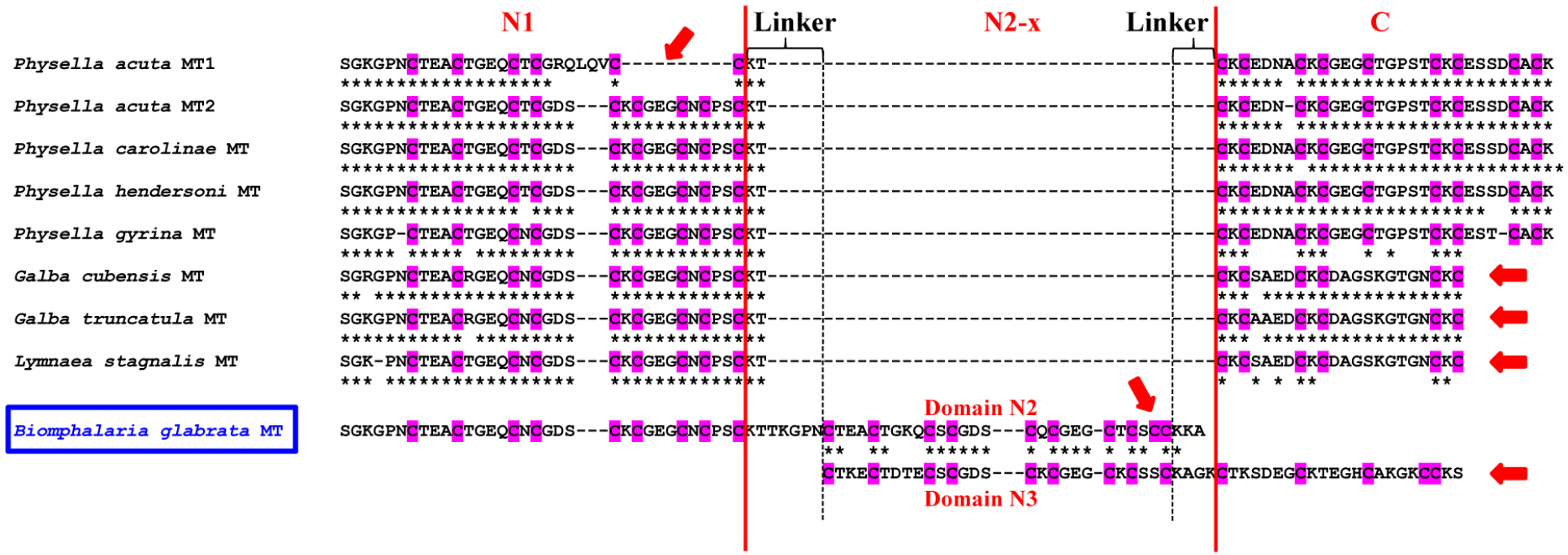
B Metal-binding domain organization and amino acid sequence alignment of Cd-selective MTs from the gastropod clades of Patellogastropoda, Caenogastropoda and Heterobranchia using the same annotations as described in Figure 1A. The bold red arrows on the right hand of the alignments points to sequence differences or similarities between MTs of the three clades.



C Metal-binding domain organization and amino acid sequence alignment of multi-domain MTs from the gastropod clades of Caenogastropoda (framed in red) and Heterobranchia (framed in blue), compared with the bidominial CdMT of *Helix pomatia* (last sequence within Heterobranchia). Numbers of Cys positions per domain are given in blue letters above the alignments. The bold red arrow on the right hand of the alignments points to sequence similarities between MT sequences of Caenogastropoda and Heterobranchia.



D Metal-binding domain organization and amino acid sequence alignment of two and multi-domain MTs from freshwater families (with Calyptraeidae, Ampullariidae and Buccinidae) of the clade of Caenogastropoda. The bold red arrows above or besides some sequences point to sequence irregularities such as truncations or Cys replacements.



E Metal-binding domain organization and amino acid sequence alignment of unspecific two and multi-domain MTs from the freshwater snail order of Hygrophila of the clade of Heterobranchia. N and C-terminal domains are designated in red letters above the alignments by N1, N2-x and C, and as Domain N2 and Domain N3 for *Biomphalaria glabrata*. The bold red arrows above or besides some sequences point to sequence irregularities such as gaps, truncations or Cys replacements.

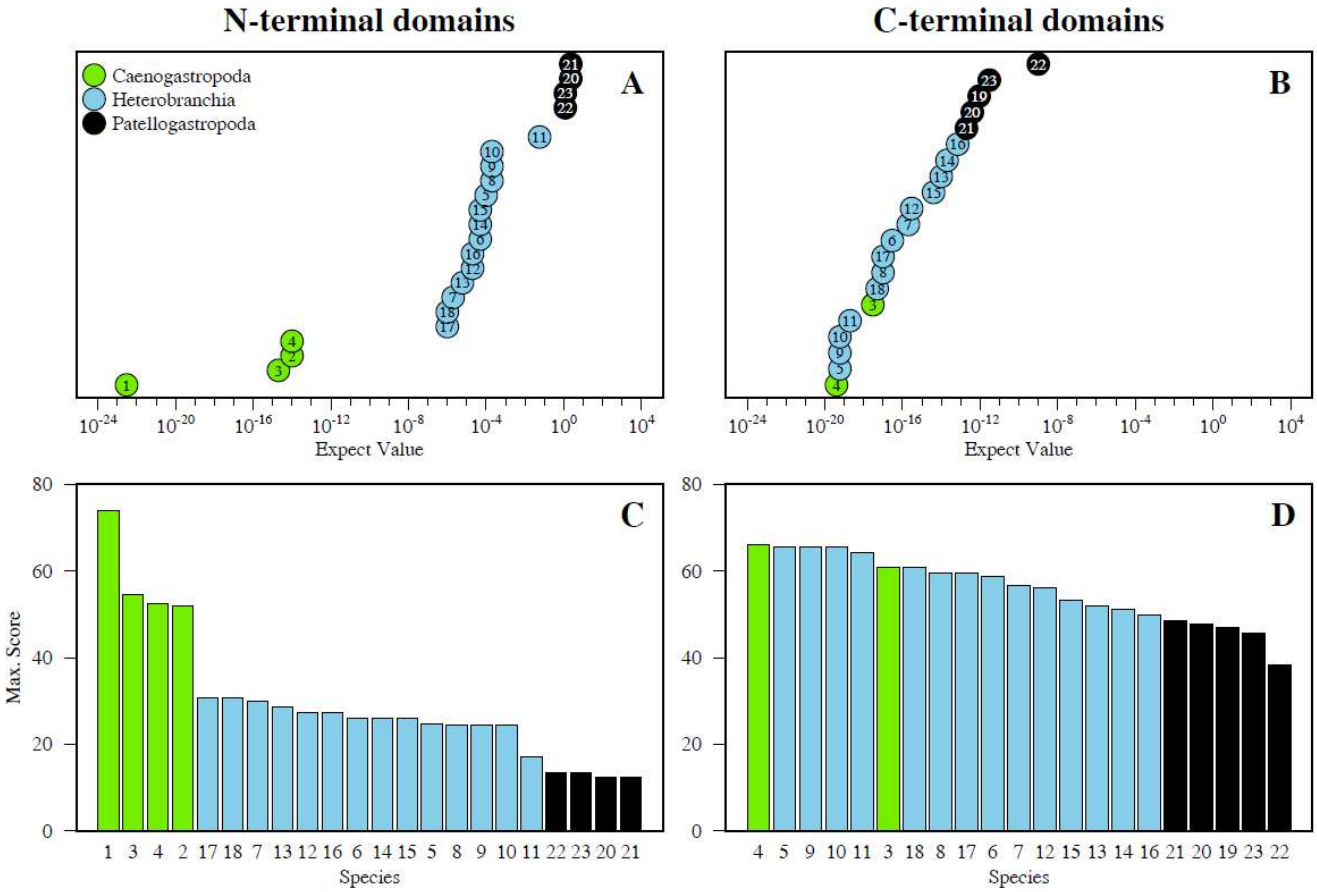


Figure 2: Conservation between N- and C-terminal domains in CdMTs of Caenogastropoda (green symbols or bars), Heterobranchia (blue symbols or bars) and Patellogastropoda (black symbols or bars). Shown are Expect Values in ascending order (**A, B**) and Homology Scores in descending order (**C, D**), calculated with BLASTp. N-terminal domains (**A, C**) and C-terminal domains (**B, D**) of Cd-selective MTs were compared to the N-terminal (N1-domain) and to the C-terminal domain of *Littorina littorea* MT. N-terminal domains are less conserved through evolution (higher differences in e-values and scores) than the C-terminal domains.
Species labels: 1, *Littorina littorea* MT (N2-domain); 2, *Pomatias elegans* MT1 (N1-domain); 3, *Pomatias elegans* MT1 (N2- & C-domain); 4, *Pomatias elegans* MT2; 5, *Helix pomatia* CdMT; 6, *Cornu aspersum* CdMT; 7, *Arianta arbustorum* CdMT; 8, *Cepaea hortensis* CdMT1; 9, *Cepaea hortensis* CdMT2; 10, *Cepaea nemoralis* CdMT; 11, *Cochlicella acuta* CdMT; 12, *Nesiohelix samarangae* CdMT; 13, *Alinda biplicata* CdMT; 14, *Deroceras reticulatum* CdMT1 15, *Deroceras reticulatum* CdMT2; 16, *Arion vulgaris* AvMT1; 17, *Lehmannia nyctelia* CdMT; 18, *Limax maximus* CdMT; 19, *Lottia gigantea* MT1; 20, *Lottia gigantea* MT2; 21, *Nacella polaris* MT; 22, *Patella vulgata* MT1; 23, *Patella vulgata* MT2.

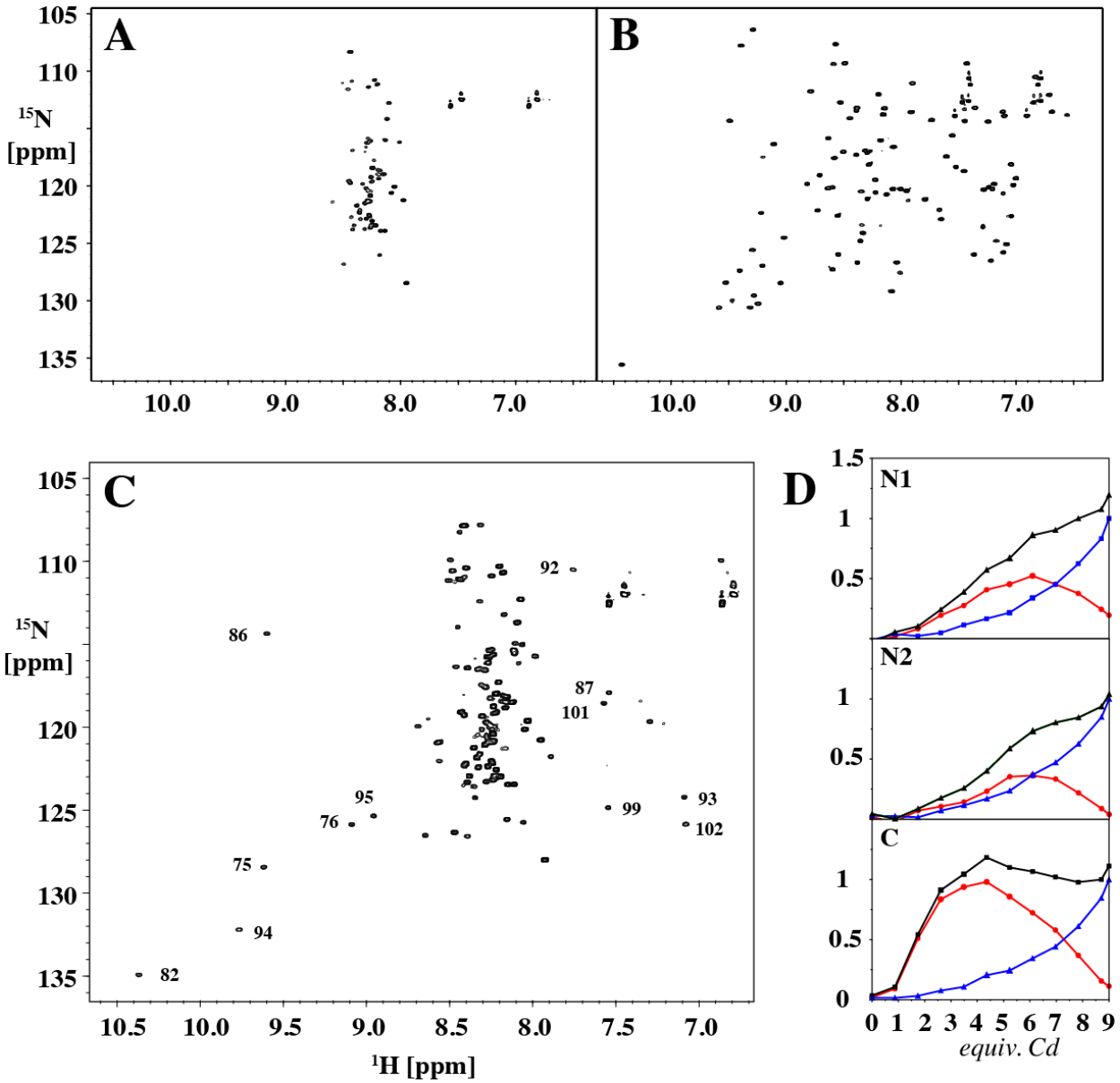


Figure 3: Metallation of apo-MT from *Littorina littorea* (LlMT) with Cd^{2+}. [^{15}N,^{1}H]-HSQC spectrum of apo (**A**) and fully-metallated (**B**) (Cd_9)-LlMT. **C** [^{15}N,^{1}H]-HSQC spectrum after addition of 2 equiv. of Cd^{2+} to apo-LlMT. Peaks close to positions in the fully-metallated form are annotated, and exclusively stem from the metallated C-terminal domain. **D** Normalized average relative peak volumes of peaks from the first (top) and second (center) N-terminal as well as the C-terminal (bottom) domains. Mostly, two peaks are observed for each residue corresponding to apo (red) and metallated (blue) neighboring domains (in the case of peaks from the C-terminal domain that corresponds to species in which the N2 domain is already metallated). The black line corresponds to the sum intensity of both peaks.

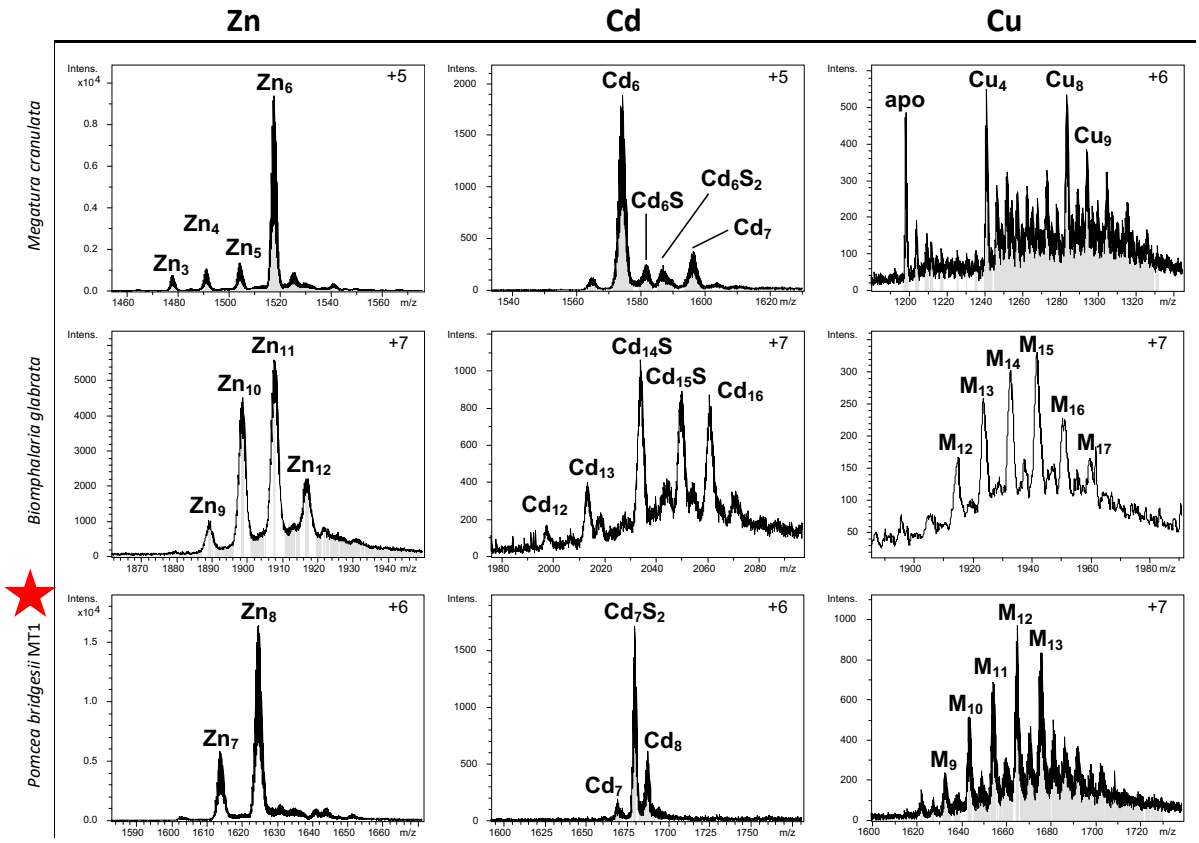


Figure 4: ESI-MS spectra of metal-unselective gastropod MTs recombinantly produced in media containing Cd, Zn and Cu ions. Data are shown for MTs from *Megathura crenulata* (Vetigastropoda), *Biomphalaria glabrata* (freshwater Heterobranchia) and *Pomacea bridgesii* (freshwater Caenogastropoda). The corresponding charge state is indicated in the upper right corner. In Cu productions, M denotes mixtures of Zn+Cu. Spectra of *pomea bridgesii* MT1 are shown here for the first time and are marked with a red star. Spectra of other MTs are re-drawn from data reported in [17,21].

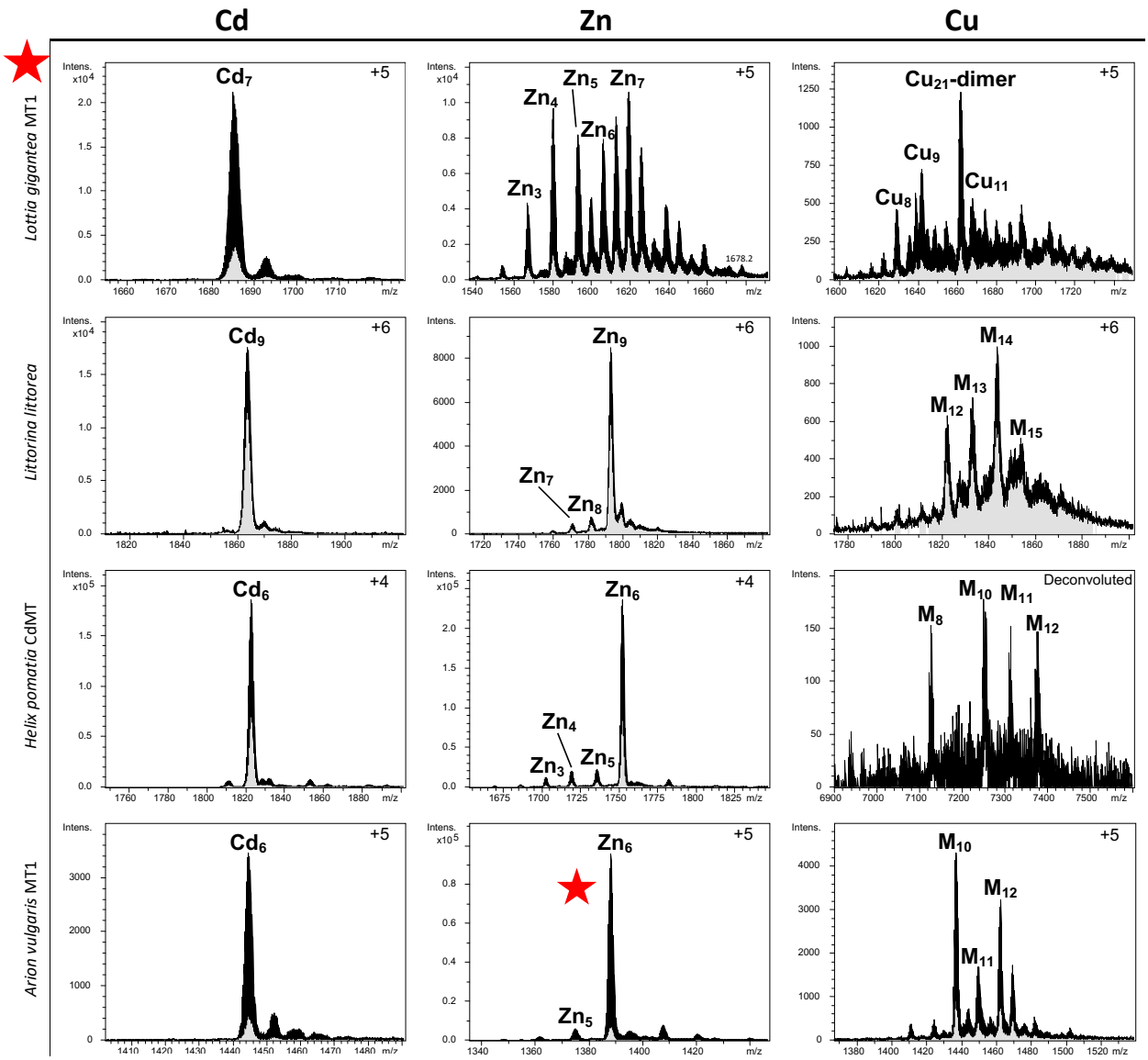


Figure 5: ESI-MS spectra of Cd-selective gastropod MTs recombinantly produced in media containing Cd, Zn and Cu ions. Data are shown for MTs from *Lottia gigantea* (Patellogastropoda), *Littorina littorea* (Caenogastropoda), *Helix pomatia* (terrestrial snail, Heterobranchia) and *Arion vulgaris* (terrestrial slug, Heterobranchia). The corresponding charge state is indicated in the upper right corner. In Cu productions, M denotes mixtures of Zn+Cu. Spectra for which metal selectivity features are shown here for the first time are marked with a red star. Spectra of other MTs are re-drawn from data reported in [22,15,23].

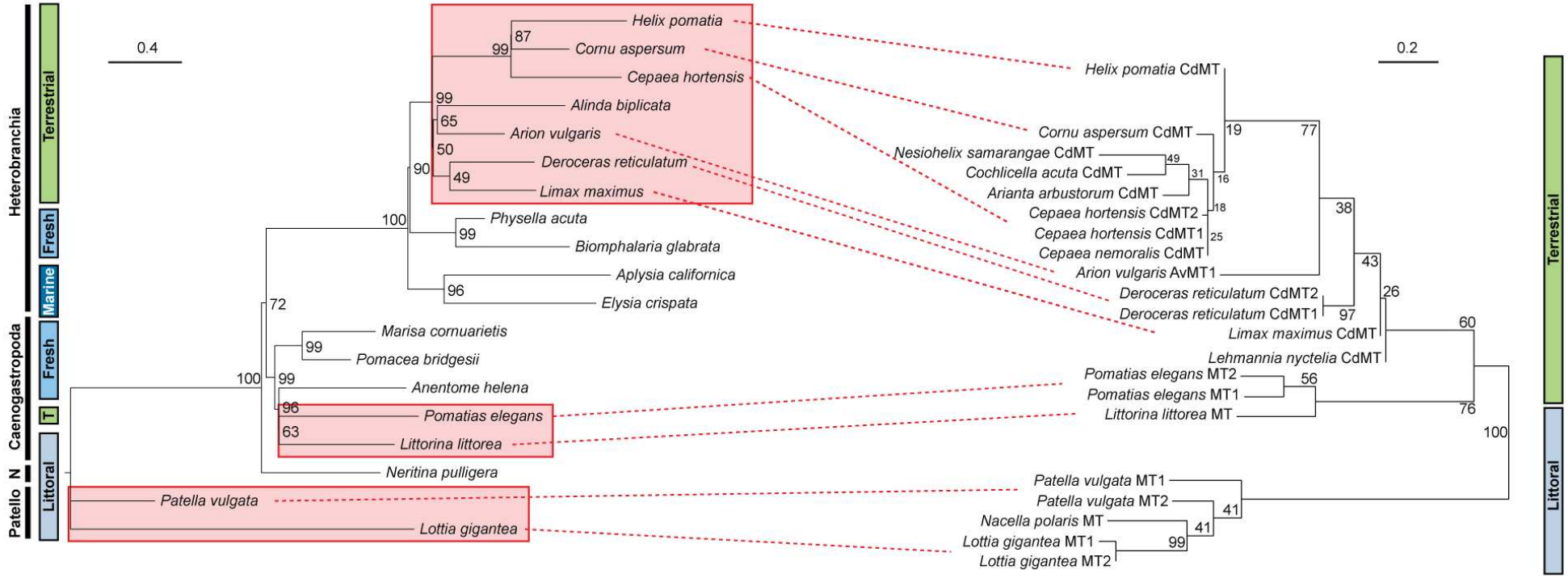


Figure 6: Mirrored phylogenetic trees (Maximum Likelihood) of investigated species of the four major gastropod clades of Patellogastropoda, Neritimorpha, Caenogastropoda and Heterobranchia. Right: phylogeny (Maximum Likelihood) showing the separated lineage clusters of only Cd-selective MTs. Bootstrap values (500 replicates) are given at nodes. Left: neutral marker phylogeny based on concatenated CO1-18SrDNA data. Bootstrap values (1000 repetitions) are given at nodes. Mirrored species possessing Cd-selective MTs are shown within red-colored frames. Identical species between the two mirroring trees are connected by dotted red lines. On outside margins of the trees, habitats of the represented species are shown with colored bars. On the left outer margin of the neutral marker tree, major taxonomic clades are indicated by black bold lines. Abbreviations: T, Terrestrial; Fresh, Freshwater; Patello, Patellogastropoda; N, Neritimorpha.

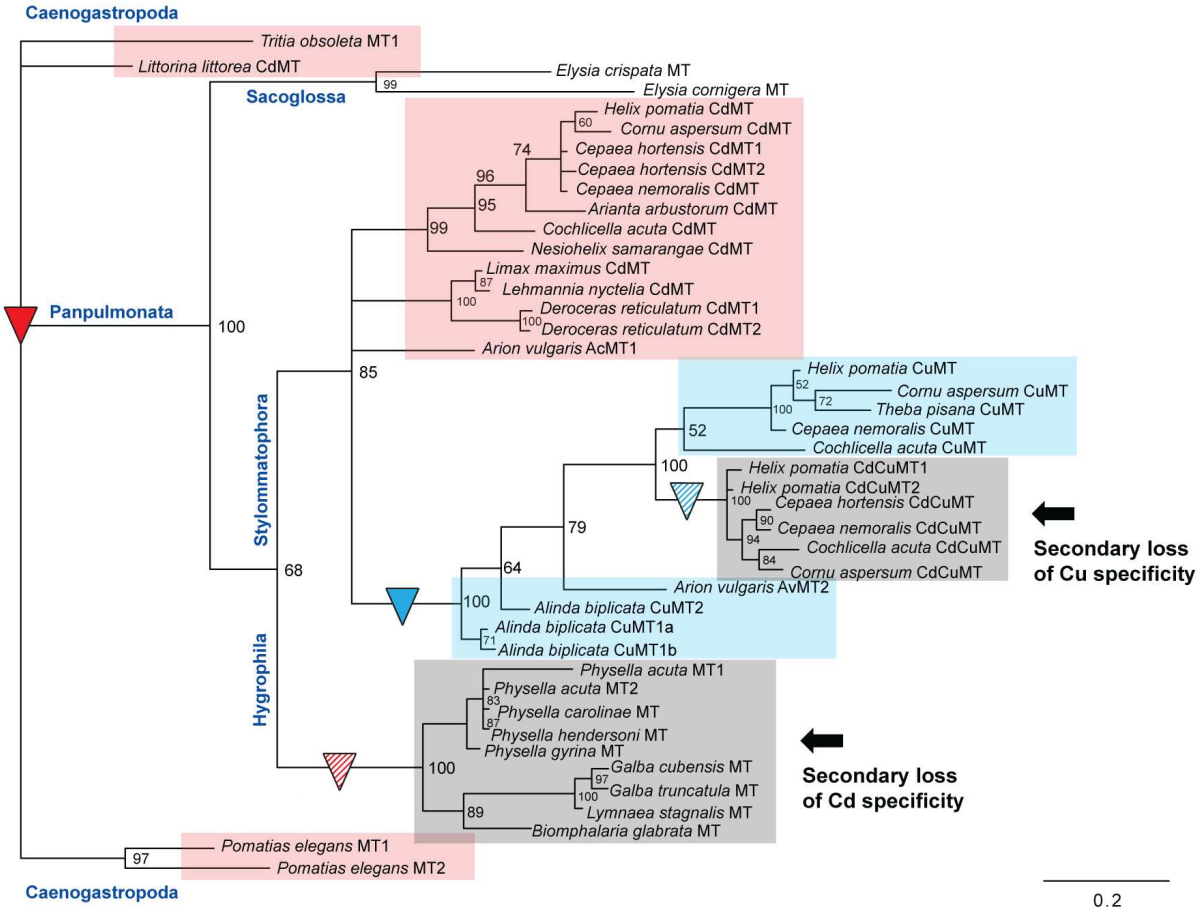


Figure 7: Bayesian Inference tree with posterior probabilities of metal selectivity features in MTs of Panpulmonata (a taxon of Heterobranchia) versus Caenogastropoda. Shown are the gain of Cd selectivity (red triangle) in MTs at the root of Caenogastropoda and Heterobranchia, with species possessing Cd-selective MTs underlaid in pink, and the gain of Cu selectivity (blue triangle) in MTs of Stylommatophora, with species possessing Cu-selective MTs underlaid in blue. Also illustrated are the secondary loss of ancestral Cd selectivity in MTs of Hygrophila (red hatched triangle), and the secondary loss of Cu selectivity (blue hatched triangle) in CdCuMTs of Stylommatophora, with respective species clusters underlaid in blue. Bayesian inference calculations were made based on a manually edited MUSCLE alignment (see **alignment S4**) using the free software MrBayes (see Material and Methods).

Figure 8

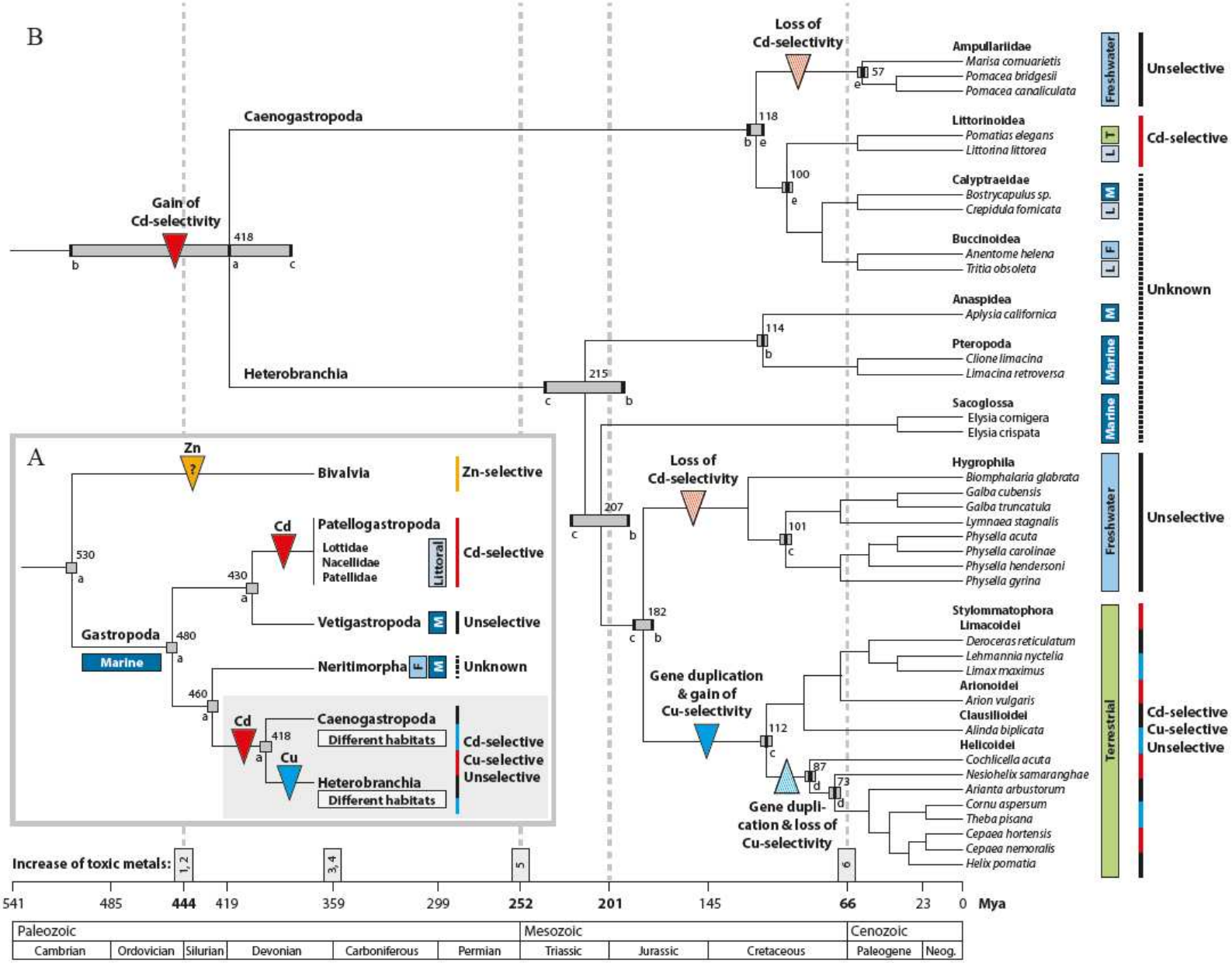

B
Loss of Cd-selectivity
Ampullariidae
Marisa cornuarietis
Pomacea bridgesii
Pomacea canaliculata
Freshwater
Unselective
Caenogastropoda
Littorinoidea
Pomatias elegans
Littorina littorea
Cd-selective
Calyptraeidae
Bostrycapulus sp.
Crepidula fornicata
Gain of Cd-selectivity
Buccinoidea
Anentome helena
Tritia obsoleta
Unknown
Anaspidea
Aplysia californica
Pteropoda
Clione limacina
Limacina retroversa
Marine
Heterobranchia
Sacoglossa
Elysia cornigera
Elysia crispata
A
Zn
Bivalvia
Zn-selective
Loss of Cd-selectivity
Hygrophila
Biomphalaria glabrata
Galba cubensis
Galba truncatula
Lymnaea stagnalis
Physella acuta
Physella carolinae
Physella hendersoni
Physella gyrina
Cd
Patellogastropoda
Lottidae
Nacellidae
Patellidae
Littoral
Cd-selective
Vetigastropoda
Unselective
Gastropoda
Marine
Neritimorpha
Unknown
Stylommatophora
Limacoidei
Deroceras reticulatum
Lehmannia nyctelia
Limax maximus
Arionoidei
Arion vulgaris
Clausilioidei
Alinda biplicata
Helicoidei
Cochlicella acuta
Nesiohelix samaranghae
Arianta arbustorum
Cornu aspersum
Theba pisana
Cepaea hortensis
Cepaea nemoralis
Helix pomatia
Gene duplication & gain of Cu-selectivity
Gene duplication & loss of Cu-selectivity
Terrestrial
Cd-selective
Cu-selective
Unselective
Caenogastropoda
Different habitats
Cd-selective
Cu-selective
Unselective
Cu
Heterobranchia
Different habitats
Increase of toxic metals:
541 485 444 419 359 299 252 201 145 66 23 0 Mya
Paleozoic
Cambrian
Ordovician
Silurian
Devonian
Carboniferous
Permian
Mesozoic
Triassic
Jurassic
Cretaceous
Cenozoic
Paleogene
Neog.

Figure 8: Phylogenetic tree (A) and chronogram (B) of Cd and Cu selectivity gain and loss in metallothioneins of Gastropoda. A (inset), phylogenetic tree of Gastropoda (reconstructed after [41,43]) with most probable relationships of gastropod clades (Patellogastropoda, Vetigastropoda, Neritimorpha, Caenogastropoda and Heterobranchia), rooted against the Gastropod sister class of Bivalvia (mussels) [92]. Gain of Cd and Cu selectivity is indicated by red and blue triangles. The kind of metal selectivity in Neritimorpha MTs is still unknown. The possible gain of Zn selectivity in Bivalvia is indicated by an orange triangle with a query. Approximate divergence times (with references) of gastropod lineages are given in million years. Marine (M), littoral (L), freshwater (F) and terrestrial (T) habitats are specified in colored framed boxes. Metal selectivities are indicated by red (Cd-selective), blue (Cu-selective), orange (Zn-selective) and black (unselective) bars. **B** Chronogram showing gains and losses of metal selectivity in MTs of the two gastropod sister clades Caenogastropoda and Heterobranchia (enhanced from grey area in **A**), with their splits into major lineages, including investigated species. Cd-selective MTs (red triangle) appeared prior to the divergence of Canogastropoda and Heterobranchia, and Cu selectivity (blue triangle) in MT isoforms of Stylommatophora. Cu selectivity was lost in novel MT isoforms of Stylommatophora (hatched blue triangle), and Cd selectivity was lost (hatched red triangles) in freshwater lineages of Ampullariidae (Caenogastropoda) and Hygrophila (Heterobranchia). Approximate divergence times of gastropod lineages are given in million years ago. Grey bars indicate published mean values for the divergence times (references a – e). Vertical, grey dashed lines indicate four of the major mass extinction events. Elevated levels of toxic metals (including Cd) are indicated in grey boxes (references 1 – 6) above the time axis. Chronogram construction was based on: a, [41]; b, [39]; c, [25]; d, [42]; e, [44]. (Additional references: [40,43,93–95]). Dating of increased volcanic Cd or metal emissions are based on information from the following studies: 1, [60]; 2, [63]; 3, [62]; 4, [58]; 5, [57]; 6, [59]. (Additional references: [2,56]).

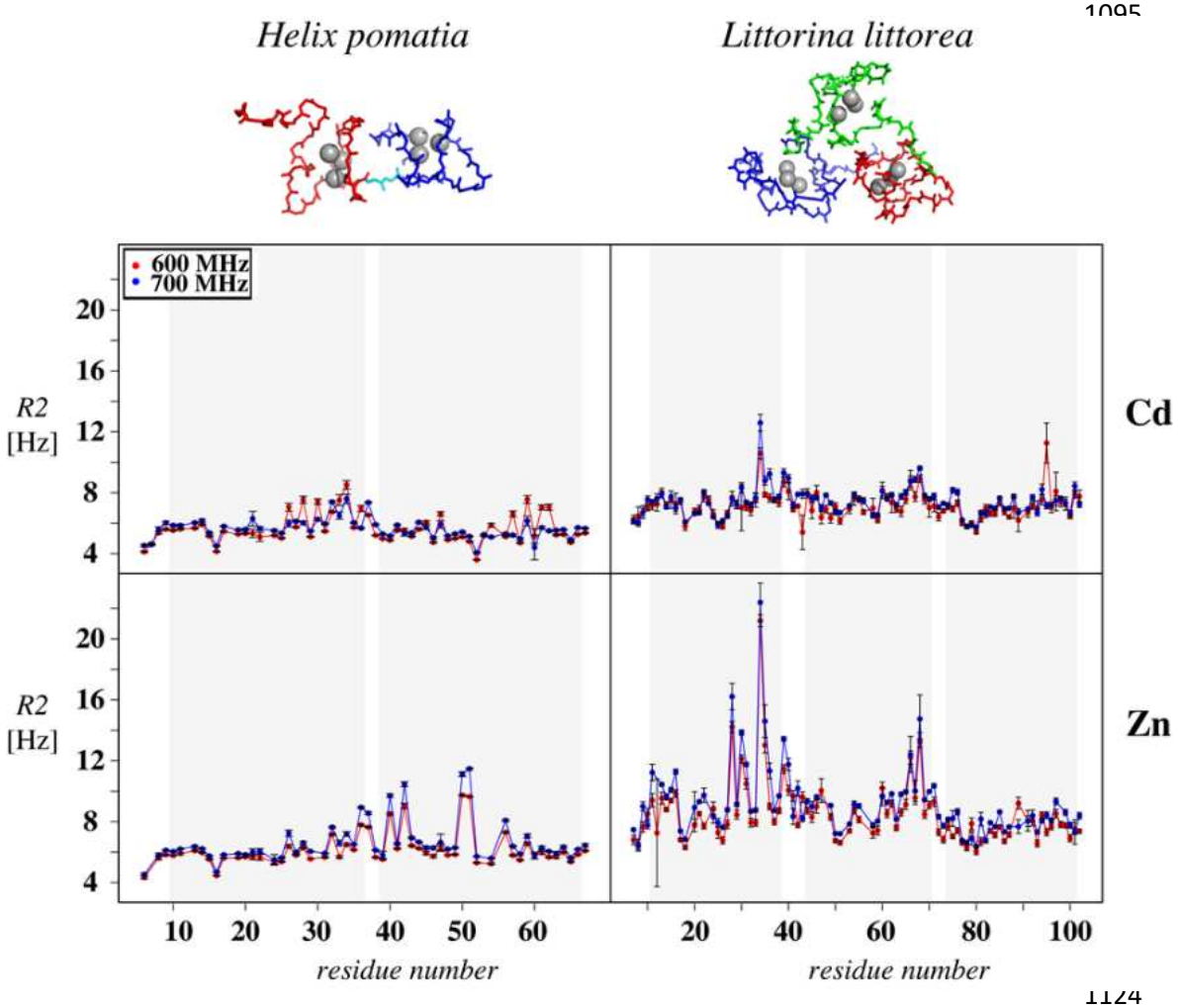


Figure 9: Conformational exchange effects in MTs. ^{15}N *R2* rates of the Cd (top) and Zn complexes (bottom) of CdMTs of *Helix pomatia* (left) and *Littorina littorea* (right), recorded at 600 (red) and 700 (blue) MHz. Contributions from conformational exchange can be detected for residues with largely increased *R2* rates.

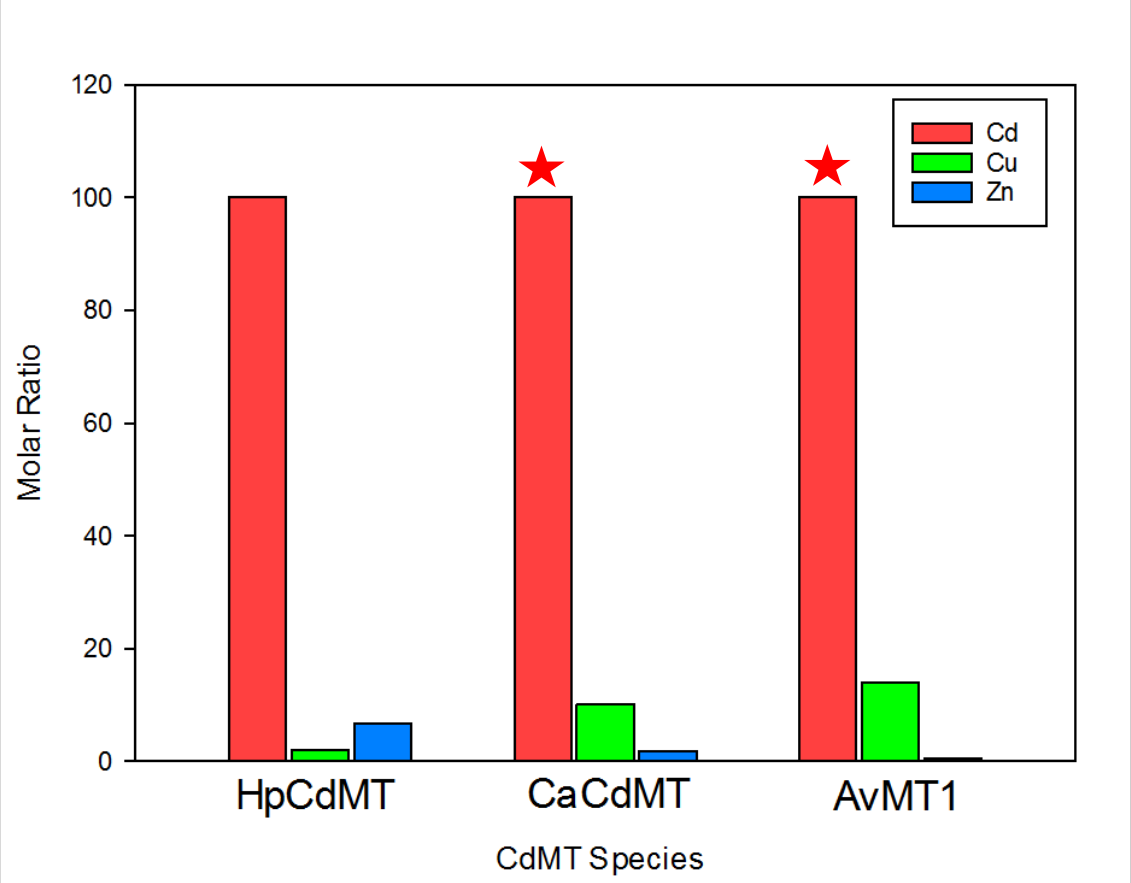


Figure 10A: **Zinc and copper content in native CdMTs isolated from midgut gland preparations of Cd-exposed snails.** Values are given as molar ratios in % of Cd content. HpCdMT, *Helix pomatia* CdMT; CaCdMT, *Cornu aspersum* CdMT; AvMT1, *Arion vulgaris* AvMT1 (CdMT). MTs of species, for which metal contents were analyzed for the first time in this study are marked with a red star. Molar ratios for HpCdMT were re-drawn from data reported in [14].

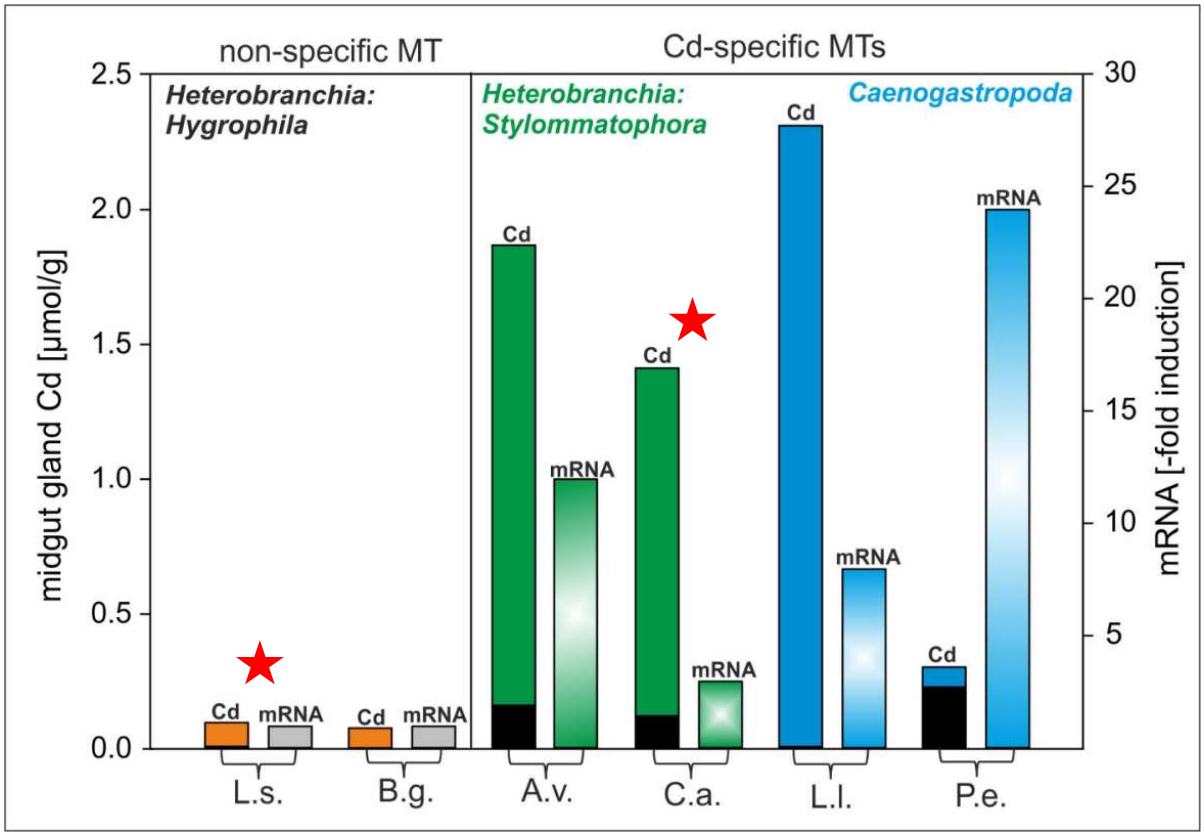


Figure 10B: Cd accumulation and –fold induction of *MT* gene mRNA transcription in midgut gland of Cd-exposed snails with unselective and with Cd-selective MTs. Left-hand part of the graph: Cd accumulation (orange bars) and –fold *MT mRNA* induction (grey bars) in two freshwater snails (*Lymnaea stagnalis* and *Biomphalaria glabrata*, both *Hygrophila*) with unspecific MTs. Right-hand part of the graph: Cd accumulation (Cd) and –fold *MT mRNA* induction for snails possessing Cd-selective MTs, with respective values for *Arion vulgaris* and *Cornu aspersum* belonging to the clade of Stylommatophora (black/green and light green bars), and for *Littorina littorea* and *Pomatias elegans* belonging to the clade of Caenogastropoda (black/blue and light blue bars). Cd contents and mRNA induction data of species analysed de novo for the present study are marked with a red star. The other values were re-drawn from data reported in [96], [23], [97] and [16]. Species abbreviations: L.s., *Lymnaea stagnalis*; B.g., *Biomphalaria glabrata*; A.v., *Arion vulgaris*; C.a., *Cornu aspersum*; L.l., *Littorina littorea*; P.e., *Pomatias elegans*.

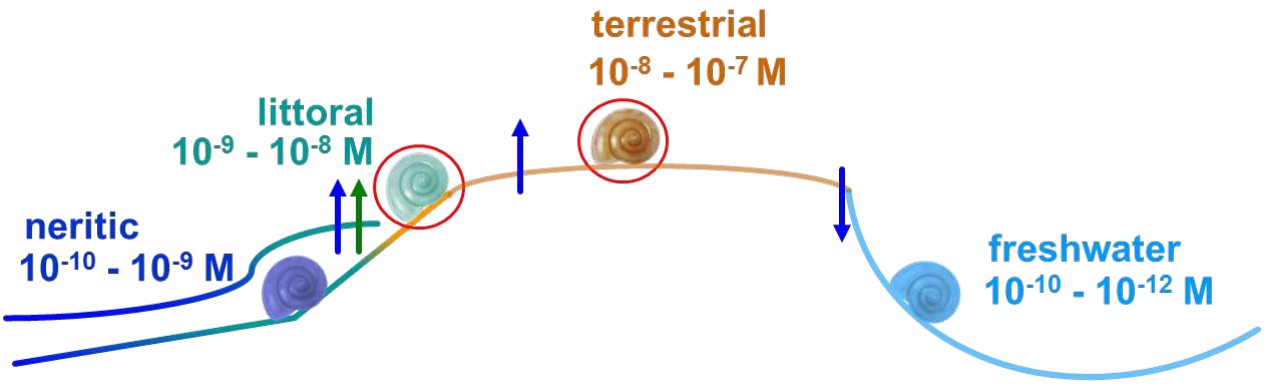


Figure 11: **Molar Cd background concentrations along the axis of evolutionary habitat adaptation in Gastropoda.** Data start from neritic superficial seawater realms, through littoral and terrestrial habitats up to freshwater environments. Blue arrows indicate an increase (upward) or a decrease (downward) of molar Cd concentrations along the habitat axis. The green upward arrow at the transition zone between neritic and littoral habitats symbolizes the increasing availability of Cd due to decreasing concentrations of complexing ligands and decreasing salinity. Snail symbols encircled in red indicate the gain of Cd-selective MTs.

Table 1 – List of gastropod species and their use for different methodical applications (red check marks) within the present study. Reported are all species acquired (first column) and their utilization for Cd exposure (second column), RNA sequencing and transcriptome assembly (second column), RNA isolation and cDNA transcription (third column), quantitative Real-Time PCR (fourth column), Protein purification from tissues *in vivo* (fifth column), recombinant expression (sixth column), MS analysis (seventh column), NMR analysis and metal titration (eighth column), and construction of neutral marker phylogeny (ninth column).

Animal collection, purchasing and rearing (*Species*)	**Cd exposure**	**RNA seq and transcriptome assembly**	**RNA isolation and cDNA**	**Quantitative RT-PCR**	***In vivo* protein purification**	**Recombinant expression**	**MS analysis**	**NMR and metal titration**	**Neutral marker phylogeny**
Lottia gigantea			√			√	√		√
Patella vulgata		√	√						√
Neritina pulligera		√	√						√
Littorina littorea						√		√	√
Pomatias elegans									√
Marisa cornuarietis			√						√
Pomacea bridgesii						√	√		√
Anentome helena		√	√						√
Aplysia californica			√						√
Elysia crispata		√							√
Physella acuta			√						√
Lymnaea stagnalis	√		√	√					
Biomphalaria glabrata									√
Arion vulgaris					√	√	√		√
Deroceras reticulatum			√						√
Limax maximus									√
Helix pomatia						√		√	√
Cepaea hortensis			√						√
Cornu aspersum	√		√	√	√				√
Alinda biplicata		√	√						√